EVOLUTIONARY AND MECHANISTIC INTERACTIONS IN VIRUS TRANSMISSION BY APHIDS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Patricia Valle Pinheiro

January 2017

EVOLUTIONARY AND MECHANISTIC INTERACTIONS IN VIRUS TRANSMISSION BY APHIDS

Patricia Valle Pinheiro, Ph. D.

Cornell University 2017

Virus transmission by aphids involves a combination of biological players that have co-evolved over years of interaction: plant hosts, aphid vectors, plant viruses and the aphid's bacterial endosymbiont. In this interaction, plant hosts activate their immune defenses against plant viruses and aphids and the latter two use a myriad of strategies to overcome, counteract or skip the host plant defense, which sometimes is done in collaboration. Bacterial endosymbionts of insect vectors might also play a role in virus transmission, either directly or indirectly, as I discussed in a critical literature review. I used proteomics and aphid genetics to show that the "biotype" phenotype of aphids is generated by genetic recombination in sexual reproduction and that aphid virulence and virus transmission are independent traits. I also showed that the host plant where aphids are reared on affects the aphid ability to transmit a circulative virus, the Luteovirid *Potato leafroll virus* (PLRV). Using organismal, biochemical, molecular, and imaging approaches, I show that the differential expression and activity of gut cysteine proteases at the cell membrane of aphids reared on a PLRV non-host plant is responsible for the change in virus transmission phenotype. Finally, using small RNA sequencing, I showed that aphids do not

activate their small interference RNA (siRNA) antiviral defense against PLRV, which provides additional evidence for the lack of replication of Luteovirids in their aphid vectors. However, aphids produce 22 nt long siRNA as an immune defense against an aphid virus, *Myzus persicae Densovirus* (MpDNV), which infects and replicates in the aphid. Strikingly, an abundance of unusually large sRNA, from 33 to 38 nt that aligned to MpDNV were produced only in aphids fed on PLRV-infected plants, suggesting that feeding on a plant infected with a circulative virus modulates the aphid antiviral immune defenses. The function of these large sRNAs is not yet known.

BIOGRAPHICAL SKETCH

I was born in the small town of Anapolis, in the state of Goias, located in the heart of the Brazilian countryside, in 1978, where I had a very nice childhood surrounded by many cousins, grandparents, uncles and aunts, an older brother and the love of my parents. My father used to be a farmer before getting married to my mother, but although he taught me many important things that I will always remember, he did not have time to teach me his love for nature and agriculture because he died of a fulminant cancer when I was 11 years old. As a child, my dream was to become a writer, but as I grew up, I became more interested in nature and biology. I went to college at the University of Goias when I was 17 and during my undergraduate studies, I worked for three years at the Entomology lab of Embrapa, our government institution for agriculture research. This was when I fell in love with Entomology and with my favorite insect, the whitefly *Bemisia tabaci*. I graduated with a Bachelor in Agronomy in 2002 and in this same year I was approved in a competitive selection process for a technician position at Embrapa, where I have worked since then. In 2003, God blessed me with my beautiful son Arthur, who gave me the strength to obtain my Master's degree in Agronomy with emphasis in Entomology by the University of Brasilia in 2005. In 2010 I applied for a scientist researcher position at Embrapa in another competitive and big selection process, based on written exams on many subjects and curriculum analysis, competing with more than 3000 scientists from all states of Brazil. I ranked in second place nationwide for the position I applied and was hired by Embrapa as a scientist researcher. In this same year I married my beloved husband Rogerio, who is my best friend and my safe haven. In the new position, I applied for an Embrapa scholarship to pursue my PhD and I chose Cornell because I was looking for a place where

I could learn new skills and mastering the ones I already had and where my creativity would be welcome. I was very happy to be accepted at the Department of Entomology at Cornell University and even happier when I started working with Dr. Michelle Cilia, who is an expert in what I was most interested in learning, insect proteomics.

Returning to Brazil and to my position at Embrapa, I plan to start a research program in insect biology using the tools I learned at Cornell to help the Plant Breeding team at Embrapa to find solutions for the management of insect pests.

To my son Arthur, who gives me inspiration and the strength to keep going, the light and joy of my life and to my new inspiration, my little baby Gabriel who is coming to our family along with the closing of this incredible journey. To my husband Rogerio, my love and my best friend, my partner in life.

“The function of education is to teach one to think intensively and to think critically. Intelligence plus character - that is the goal of true education.” Martin Luther King Jr.

If you can’t explain it simply, you don’t understand it well enough.” Albert Einstein

“Be kind whenever possible. It is always possible.” Dalai Lama

“Faith, trust and a little bit of pixie dust!” Peter Pan

ACKNOWLEDGMENTS

I am grateful to God for all the nice people that I met during this rich spiritual and cultural journey, for everything that I have learned in science and for all my accomplishments. I have been blessed with so many people that helped me during this joyful journey.

I am especially grateful to three people who believed I could do it: my husband Rogerio Silva dos Santos, my advisor Michelle Cilia and my research leader at Embrapa and mentor, Flavio Breseghello.

To my dear husband and best friend Rogerio for all his love, care, support and patience in our relationship. For helping me with everything, including experiments, parenting, taking care of our house and of our life and of my mental health, and for always being there for me.

To my dear friend and advisor Michelle, who saved me from quitting my PhD and who taught me everything with love and patience, I will always be grateful.

To my mentor Flavio, one of the best scientists I have ever met, for suggesting me to applying to Cornell (and for believing that I could be accepted at Cornell, while I myself found it hard to believe).

I will be always grateful to my son Arthur for his understanding and love in all the times I could not be with him or give him the support and care he deserves. For being my safe haven and for helping me understanding the real meaning of life.

To my father Ronan, in memory, for his love and dedication during the 11 years we lived together.

To my mom Helvecia and my brother Leonardo, for giving me their unconditional support at all moments of my life. To my mom, I will always be grateful for her dedication, for taking care of me and my brother and for showing us, by giving the example, how to become honest and honorable people. I will always be grateful for everything that she did for me and for Arthur and for always being there for us. To my brother Leonardo, who provided for me and for my mom

while I was completing my undergraduate studies. Without them I would never have accomplished so much. Also, for taking care of my businesses in Brazil while I was away, for visiting us so many times, giving us the strength to stay so far from home.

To my family in Brazil, my aunts Neide and Fatima, my cousins Fabiana, Carolina, Michelle and Priscilla for their support, for letting me know that I could count on them, for taking care of my mom and of our grandpa, while I was away.

To Gary Blissard, Hudson Kern Reeve, Stewart Gray and Zhangjun Fei (alphabetical order), for serving on my thesis committee and for helping me with precious comments and suggestions on my research and dissertation.

To all members of the Cilia lab, I cannot thank you enough and I will be always grateful to each and every one of you for the joyful moments we spent together. Thank you for always being nice and helpful, for making the lab a nice and collaborative environment.

To Stewart Gray for all the support I received in his lab with my experiments, from the use of his lab, growth chambers, greenhouses, reagents, equipment and time of his lab members, for all his advice, help in the discussion, for sharing his valuable experience and for being patient with me.

To Dawn Smith for caring of aphids and plants, for always being helpful and nice, for discussing the design of experiments and sharing her precious experience with kindness and care.

To Jason Ingram for all the support he gave me in my experiments, for his patience with my uncountable requests and for fumigating hundreds of plants from my experiments.

To Kevin Myers for being kind and providing me with turnip plants every time I needed.

To Yi Xu for sharing his knowledge, kindness and support and for being a nice friend.

To Margarita Lopez Uribe for helping me in a desperate moment.

To my GCAs colleagues at Hasbrouck for all the learning, fun and work that we

shared. To my supervisor at Hasbrouck, Linda Majani, who taught how to be in peace with things I can't change.

To Surya Saha, Yi Zheng, Wenbo Chen and Noe Fernandez for their help with bioinformatics analysis, discussion and training.

To Mamta Srivastava for her help in the Plant Cell Imaging Center at BTI and for being nice and patient at all the times that I had to repeat the same questions about the microscopes.

To Murad Ghanim for helping me obtaining beautiful confocal images of aphids, for always being nice and for sharing his knowledge with me.

To Elaine Vanetten and to Dale Barrows for their kind support at all the times my computer died or was broken.

To my dear friends in Ithaca, who made my days here happier: Isis Caetano, Kristen Brochu, Vanessa D'Absalon and Vitor Medina, Arkangel and Irma Cordero, Sergio Ospina and Martha Mateus, Lorena Carvalho and Marcus Passarinho, Jaime and Paula Ortiz and Mia Greenwood for their friendship, for listening to me, for discussing science and other (more) important things, for their love and for their time.

To all the staff of the Cornell Fitness Centers, where I got some relief from the stressful routine as a PhD student.

To the staff of the Cornell library system, my favorite part of Cornell and the part I will miss the most.

To Embrapa, NSF and Cornell for providing funding resources for my scholarship, research and travel funds to conferences.

TABLE OF CONTENTS

Biographical sketch…… ……..…..…..iii
Dedication……..v
Acknowledgements……..……..vi
Table of Contents……..…….ix

Introduction……1

Chapter 1. Evidence of the biochemical basis of host virulence in the greenbug aphid, *Schizaphis graminum* (Homoptera: Aphididae)…….……45

Chapter 2. Increased lysosomal exocytosis in the aphid midgut decreases circulative virus transmission……99

Chapter 3. Is there a role for symbiotic bacteria in plant virus transmission by insects?.....162

Chapter 4. Evidence for aphid immune system modulation by plants infected with circulative viruses but not stylet borne viruses……195

Appendices…….……..245

I) The conditional vectoring hypothesis……..246
II) The draft genome of the whitefly *Bemisia tabaci*……248

Introduction

Are aphids, Buchnera and Luteovirids a superorganism?

Review article to be submitted to the IPVE Symposium Special Issue of Virus Research in December 2016

Patricia V Pinheiro [1,2,3] H. Kern Reeve[4] and Michelle Cilia[1,5]

1 Department of Entomology, Cornell University, Ithaca, NY
2 Boyce Thompson Institute for Plant Research, Ithaca, NY
3 Embrapa Rice and Beans, Santo Antônio de Goiás, Brazil
4 Department of Neurobiology and Behavior, Cornell University, Ithaca, NY
5 USDA Agricultural Research Service and Plant Pathology and Plant Microbe Biology Section at Cornell University, Ithaca, NY

Virus transmission by insect vectors involves a combination of biological players that have co-evolved over years of interaction: plant hosts, aphid vectors, plant viruses and the aphid bacterial endosymbiont. In this system, plant hosts activate their immune defenses against plant viruses and aphids and the latter two use a myriad of strategies to overcome, counteract or skip the host plant defenses (1–3), which sometimes is done in collaboration (4–7). Bacterial endosymbionts of insect vectors might also play a role in virus transmission, either directly or indirectly (8).

Aphids and their bacterial endosymbiont, *Buchnera aphidicola*, depend on each

other for survival. Plant viruses in the family Luteoviridae, called here as Luteovirids, are an interesting example in the vector biology field because these viruses will infect only plants that are in the host range of their aphid vector. On the other hand, it has also been shown that aphids are more attracted to plants infected with viruses they transmit than to healthy plants or to plants infected with viruses that have different modes of transmission (9, 10). For some systems, the progeny and survival of aphids fed on Luteovirid-infected plants is increased, compared to aphids fed on healthy plants (11), which suggests that feeding on a Luteovirid-infected plant benefits the aphid vector. Altogether, these results suggest that aphids, their bacterial endosymbiont and Luteovirids are a superorganism. In this scenario, plant hosts of aphids and Luteovirids are a target of a superorganism that act together to skip plant defenses and therefore, these plants need to mount a diversified range of strategies as a response to aphids and Luteovirids. In this review we will look at the current knowledge of the vector biology field on the aphid-Luteovirid interactions and propose a model where aphids, endosymbionts and Luteovirids interact as a superorganism.

Aphids

Aphids are small insects from the order Hemiptera, with more than 4000 species spread worldwide (12). These insects feed exclusively on the phloem sap of plants, stealthily inserting their piercing-sucking mouthparts in the plant without direct contact with the cell wall and then causing little or imperceptible damage to the plant cells individually (reviewed in (13). However, due to their ability to switch from sexual to

parthenogenetic reproduction, aphid populations can increase at extraordinary rates. In great numbers, they can cause significant damage to plants by feeding and sucking nutrients from the plant and most importantly, by transmitting plant viruses. Therefore, they are considered as important pests of agronomic crops, causing significant yield losses in many areas of the world. For most of the aphid-transmissible viruses, the management of the plant disease in the field is focused on largely ineffective application of insecticides to reduce aphid populations. However, controlling aphid vectors using pesticides is costly, and to be effective, information about vector phenology is necessary. Disrupting an aphid's ability to transmit a virus into or within a crop represents a different approach and a promising means by which to control virus spread (14, 15).

Aphids are the most important and wide spread vectors of plant viruses, responsible for the transmission of more than 50% of all plant viruses transmitted by insects (16, 17). Most aphid species that transmit viruses belong to the family Aphididae (Hemiptera) (12). These insects are a perfect vehicle for virus transmission due to their feeding style, which allows them to ingest virions from an infected plant and to deliver these virions directly into living cells of a healthy plant (18). Also, their ability to reproduce parthenogenetically facilitates a rapid increase of the progeny, which promotes rapid dispersal of the aphid population and increases the fitness of the plant virus they are carrying. Phenotypic plasticity allows aphid populations to produce winged individuals when the food resources are scarce and to search for a new host plant where they can start a new colonization. This characteristic also contributes to their efficiency as vectors of plant viruses, promoting plant-to-plant spread of the viruses they transmit. It is interesting to note that for plant viruses that depend on an insect vector for transmission,

the ultimate host range is determined by the host range of the insect vector, as in the case of Luteovirids. For example, the specialist aphid *Schizaphis graminum* feeds only on a few species within a single plant family, the Poaceae, and are able to transmit multiple species of Barley and Cereal Yellow Dwarf viruses (B/CYDV). In contrast, the generalist *Myzus persicae* feeds on a wide range of plant species, from more than 40 families (12) and also transmit multiple plant viruses. Recently, we showed that the host plant where aphids are reared on has an impact on the ability of generalist aphids to transmit circulative viruses (19).

Morphologically indistinct populations of insects that present a unique relationship with a resistant host plant cultivar have been classified as “biotypes”, without considering the insect genetic background. In *S. graminum*, the classification is based upon the resistance/susceptibility of small grains cultivars (wheat, rye, barley), inbred for different resistance genes, to the new aphid population (20, 21). However, coevolutionary races between hosts and pests will typically yield genetic variability in a two-way direction: in this case, in the plant defense strategies and in the counter-defense strategy used by the insect. Genetic variability among these biotypes have been shown for the transmission efficiency of B/CYDV, for example. Not all populations (“biotypes”) of *S. graminum* are efficient vectors of B/CYDV, with vectoring efficiencies varying from 0 to 100% in a continuum (22–27). The ability of *S. graminum* to transmit B/CYDV is controlled by a few genes and protein isoforms in an additive manner (22, 23, 25, 28), which makes some biotypes and genotypes of *S. graminum* more efficient vectors of B/CYDV than others. One of these biotypes, named biotype H, presented interesting characteristics for the study of the biochemical basis of biotype virulence in aphids, such as a high efficiency of

transmission of five species of yellow dwarf viruses (29) as compared to other biotypes which transmit only one or two yellow dwarf species and the highest virulence to agronomic crops among the biotypes studied (20). Comparing biotype H of *S. graminum* to other biotypes of this aphid species with a series of proteomics, fitness and virus transmission experiments, the biochemical basis of the host virulence phenotype in aphids was studied and proved to not be controlled by the genes involved in the virus transmission phenotype (30).

Another interesting aspect of aphid biology is that almost all of the 4000 species of aphids harbor an obligate endosymbiont, the y-proteobacterium, *B. aphidicola*, in specialized cells called bacteriocytes. Symbiosis between bacterial endosymbionts and Hemipteran insects are exemplary of symbioses in nature. The relationship between aphids and *B. aphidicola* is one of the best-studied cases of symbiosis in the plant vector biology field (recently reviewed in (31). The bacteria provide amino acids that the aphid is unable to synthesize or to obtain from its plant sap diet. Both organisms depend on each other for survival: the bacteria are not culturable (32) and aphids treated with antibiotics to eliminate the bacteria grow poorly and do not reproduce well (33). A proteomics study of the pea aphid *Acyrthosiphon pisum* bacteriocyte showed the expression of metabolite transport proteins and all enzymes involved in amino acid metabolism annotated in the *Buchnera* genome (34), providing support for the nutritional function of the symbiosis for aphids. The bacteria are passed to the progeny by transovarial transfer (35, 36). The molecular mechanisms for transovarial transfer of symbiotic bacteria in insects are not known but recent work in drosophila shows the vertically transmitted *Wolbachia* spp. is controlled via the insect's diet (37).

The pea aphid *A. pisum* is the only aphid genome sequence available to date (38), although sequencing projects of other aphid species are underway. Immune defense pathways involved in the defense against bacterial infection in other insect species, such as the IMD pathway, antimicrobial peptides and peptidoglycan recognition proteins, are lacking in aphids (38, 39). It has been proposed that the deletion of entire immune pathways occurred adaptively to support the obligate symbiosis with *Buchnera*. Supporting this hypothesis, *Buchnera* cells obtained by cell fractionation of aphid homogenates and added to a culture of non-host cells (Drosophila S2 cells) were eliminated in two days and induced the expression of genes for antimicrobial peptides via the IMD pathway in the S2 cells (40). As in other endosymbiont genomes, *Buchnera* has lost most of the genes for transcriptional regulation that are present in free-living relatives, such as *Escherichia coli* (41). *Buchnera*'s genome has most genes for the synthesis of essential amino acids, but some are missing, as well as most genes for the synthesis of nonessential amino acids (42). Differential expression of genes in amino acid biosynthesis pathways between bacteriocytes and other aphid tissues indicate complementarity between amino acid pathways encoded by the host and symbiont genomes (42). Transcription in *Buchnera* is stable, with no detectable differential expression of mRNA (43). Other mechanisms for regulation of gene expression in *Buchnera* have been proposed, such as small RNAs and regulated protein stability (43). Altogether, these data provide evidence for the reduction of the aphid immune system as an evolutionary adaptation to its obligate endosymbiont, as well as the reduction of the *Buchnera*'s gene repertoire for transcriptional regulation.

Host plants of aphids

Host plant selection by aphids is a complex process that involves the recognition of physical and chemical cues (for a review, please see (44). Attracted by visual cues and plant volatiles, winged aphids land on a plant and start the assessment of plant surface, which includes stylet penetration for probing. Before establishing a long term feeding, aphids initiate a few short-term stylet penetrations to probe the plant epidermis. When probing a suitable host plant, aphids penetrate the mesophyll and parenchyma tissues with their stylets until they reach a sieve element. Once in the sieve element, feeding is accompanied by salivation and injection of watery saliva. A prolonged period of feeding of ten minutes or longer may represent host plant acceptance. Aphids can keep a sustained feeding on a host plant for several hours (45). Therefore host acceptance depends on stylet penetration and sustained feeding (45).

Aphids colonize a wide range of host plants. The mechanisms that determine host range in aphids are not well understood. For example, *S. graminum* can feed and colonize small grain species in the Poaceae, such as wheat, rye and barley, but it cannot feed on maize (*Zea mays*), from the same family. When offered maize as their only choice *S. graminum* die in a few hours with their stylets inserted on the leaves, suggesting that these aphids probed the plant and were not able to detoxify chemical defenses produced by the plant.

In spite of the host plant range being different for specialists and generalists, aphids of both types of feeding habits have preferred hosts, on which they develop faster and produce larger individuals. Sympatric populations of the pea aphid (*A. pisum*)

complex associated with different host plant races have been reported (46). Feeding on non-preferred hosts requires a period of adaptation until the aphid population can fully colonize the plant. For example, for the cereal aphid *Sitobion avenae*, a specialist aphid that feeds on a few plant species of the Poaceae, the preferred host is wheat, but it can also colonize barley and other grasses. When a clone of *S. avenae* collected in wheat was transferred to barley, a negative host switch effect was observed, impacting life table parameters (47). Similarly, a pea host race clone of *A. pisum* reared on non-preferred hosts, exhibited small size, low reproduction rate, slow population increase and developmental rate (48). These studies support the hypothesis that specialist aphids that have preferred hosts perform relatively poorly on alternate hosts (48, 49). This kind of observation is useful to establish cultural control measures to reduce aphid populations, such as intercropping wheat with barley.

On the other hand, the ability of generalist aphids to feed and colonize a large range of plant species requires high adaptability to overcome a variety of host plant defense mechanisms from different plant species, usually by delivering effector proteins during salivation (13, 50, 51). The high phenotypic plasticity of individuals and genetic variability of populations may contribute to the polyphagy of generalist aphids, such as *M. persicae* (52). In this species a wide intraspecific genetic variability has been correlated with color (53), life cycle (54, 55), karyotype (56) and host plant adaptation (52, 57, 58). The host plant switch effect was studied in *M. persicae* by transferring aphids reared on broad beans (*Vicia faba*) to rape (*Brassica napus*) or potato (*Solanum tuberosum*), using a 2D polyacrylamide gel electrophoresis coupled with Mass spectrometry (MS) (59). These plants belong to three different families and present different defense strategies

against aphid feeding. Among their findings, the authors found some proteins involved in glycolysis, the TCA cycle and mitochondrial electron transport to be down-regulated in aphids fed on rape and potato, compared to beans, suggesting that switching hosts had a negative impact on energy allocations and possibly aphid fitness, although fitness assays were not performed. Omics approaches are very useful to investigate host switch effects on insects as well as any other phenotype of interest, but they are more informative if the results can be related to a certain phenotype. For example, proteins in the TCA cycle are upregulated in *S. graminum* biotype H, and this was correlated with higher fitness and virulence in this biotype, compared to other biotypes (30).

Since aphids are capable of efficient adaptability to new host plants, host plant adaptation might result in differences in virus transmission efficiency of an aphid population, for a particular virus species (60). Evidence supporting the idea that an aphid's host plant influences the luteovirid transmission was reported for the specialist aphid *Rhopalosiphum padi* in a field study conducted 20 years ago, but their data were not discussed in the context of a host switch effect on virus transmission (61). The authors observed BYDV-RMV to be efficiently transmitted by *R. padi* collected from corn in the field correlating to high levels of this virus moving from corn to wheat crops. However, when *R. padi* were brought into the lab and reared on barley, they were not efficient vectors of RMV. Intriguingly, an increase in temperature improved the ability of barley-reared *R. padi* to transmit RMV, suggesting that altering the activity of an aphid enzyme improved virus acquisition/transmission. Additionally, it is known that the ability of *S. graminum* biotypes to transmit B/CYDV is correlated with host plant adaptation (29). For example, *S. graminum* biotypes that were isolated from wild grasses are more efficient in

transmitting B/CYDV species than those adapted to agronomic crops.

The plant where the generalist aphid *M. persicae* is reared on has a profound effect on the aphid's ability to transmit both persistent and nonpersistent viruses, as it has been widely reported (60, 62, 63). For example, a difference in virus transmission was observed for *Zucchini Yellow Mosaic Virus* (ZYMV), a virus that is nonpersistently transmitted by *M. persicae,* when aphids were reared on two plants that are both hosts of *M. persicae* but non-hosts of ZYMV, the Brassica mustard (*Brassica juncea*) and the Malvaceous okra (*Abelmoschus esculentus*) (60). Overall, aphids reared on mustard had higher transmission rates of ZYMV to a host recipient plant (*Cucurbita pepo*) than okra-reared aphids. Interestingly, the host switch had a complex effect, giving intermediate transmission rates when mustard-reared aphids were given a preacquisition time of 24 hours on okra plants, showing that the host switch effect was transient and along a continuum. In a previous study with PLRV, 75% of the nymphs born on physalis transmitted PLRV to recipient plants, while only 49% of the nymphs born on rape (*Brassica rape*) transmitted the virus (63). It is interesting to note that physalis is a host plant of PLRV, while rape is a nonhost of this virus. A similar effect was observed in another study, when *M. persicae* reared on rape were less efficient in transmitting *Beet Yellows Virus* (BYV) than aphids reared on beet, a host of the virus (62). Again, in this case, rape is a nonhost of BYV. Therefore, it seems that rearing aphids on a host plant of the virus increases the efficiency of virus transmission by this generalist aphid, as we recently showed (19). Switching hosts from physalis to turnip, the ability of *M. persicae* to transmit PLRV was significantly reduced and the host switch effect was transient. We also provide a series of organismal, biochemical, molecular, and imaging approaches to

discuss the mechanisms involved in the host plant switch effect on virus transmission by aphids (19).

Transmission of Luteovirids by aphids

Plant viruses from the family Luteoviridae, referred to here as Luteovirids, are transmitted exclusively by aphids in a circulative manner, which means that these viruses have to circulate through the aphid tissues and cross multiple barriers until reaching the salivary duct, from where the virus particles can be ejected into a new plant during feeding and salivation (For a recent review, please see (14). Luteovirids are retained in the aphid vector for its entire life, so they are classified as persistently transmitted viruses. Interestingly, although limited replication of luteovirids in aphids has been suggested (64), there is no evidence that Luteovirids replicate in the aphid vector tissues at sufficient levels for the aphid to be considered as a host of Luteovirids. Instead, aphids are considered as a passive vehicle for virus transmission and are thought to benefit from transmitting plant viruses (5, 7, 11), which are kept at low and harmless levels in aphid tissues. However, the mechanism used by aphids to avoid the replication of Luteovirids is not known.

Luteovirids are nonspecifically ingested from the phloem sap together with sap proteins (65) while the aphid is feeding on an infected plant. To be transmitted to a new plant, Luteovirids must overcome physical barriers within the insect, the gut and the accessory salivary glands, a process that is mediated by virus-vector species-specific protein interactions. The virus must first be internalized by gut cells (65–71). Aphids

acquire and transmit Luteovirids as intact virions, not viral RNA. Detailed microscopic investigations revealed that the virus moves via endosomes in the aphid gut, with different virus species displaying different affinities to various regions of the gut (i.e., midgut or hindgut). PLRV is acquired into midgut epithelial cells (72). Virions bind to the luminal (apical) plasma membrane, stimulating the formation of coated pits and enter the gut epithelial cells via a receptor-mediated endocytosis mechanism (73). Recently, the first Luteovirid receptor in the aphid vector has been identified, the aphid membrane alanyl aminopeptidase N, a cell surface receptor for *Pea enation mosaic virus* (PEMV, genus *Enamovirus*) in the pea aphid, *Acyrthosiphon pisum* (74, 75). Once inside the cell, the virus particles remain in membrane-bound vesicles during transport through the cytoplasm, and this is universally true for every species of Luteovirid studied by microscopy to date. Unlike in plant cells (76), virions are never observed free in the aphid cytoplasm. The cytoplasm is the site of virus replication in plant phloem cells (76) where virions have been observed to decorate the mitochondrial, vacuole, and chloroplast membranes. Such isolation from the aphid cytoplasm within membrane-lined vesicles may explain why there is no virus replication in the insect. The observation that virus-containing tubular vesicles connect to aphid cellular organelles led to the hypothesis that the virus is transported intracellularly through the gut endomembrane system. Membrane-bound vesicles containing virions in gut cells of *M. persicae* and other aphid species have been observed to connect to lysosomes and lysosomal-like organelles (69, 72). Following transport through the endosome, PLRV and other Luteovirids can be observed between the plasmalemma and the basal lamina of the gut epithelia where they are then released into the open circulatory system of the aphid and quickly diffuse (69, 72). Once the virus

reaches the accessory salivary glands, the virus is endocytosed (77, 78), transported through the cells in vesicles, and released into the salivary duct where it can be inoculated into plants together with the saliva as the insects feed. Once aphids acquire a Luteovirid from an infected plant, the aphid remains viruliferous for its entire life (64). We recently provided evidence for altered activity of cysteine proteases in the aphid gut to be involved in a change in the virus transmission phenotype (19).

Luteovirids are icosahedral, positive sense RNA viruses. Luteovirids share a conserved arrangement of three open reading frames in the 3' half of their genome, two of which encode the structural proteins (79).The capsid consists of the coat protein (CP) and a minor amount of the readthrough protein (RTP) translated via a readthrough of the CP stop codon (80–83). The RTP is not required for particle assembly or infection (82–84), but particles containing only the CP are not transmissible by aphids to plants (80, 85–87). These RTP-minus virions can pass from the gut to hemocoel, but not hemocoel to salivary gland indicating that the CP alone contains the topologies required for virus-gut interactions (68, 86, 87). RTP-minus virions do not move as efficiently across the gut as wild-type (WT) virus (87). Therefore, although the RTP is not required for the virus to move across the gut membrane, it may facilitate virus uptake. Similarly, the RTP may not be absolutely required for virus to cross into the salivary gland (68), but it may improve uptake efficiency. Additionally, both of these proteins regulate virus movement in plants. The CP is required for local and systemic movement; the RTP acts *in trans* to retain virus in the phloem where it is available to aphids (88). How these two virus proteins regulate all these different activities in aphids and plants is unknown, but it is likely that virions regulate their trafficking via interactions with host and vector proteins.

Aphid populations within a species vary in their ability to transmit a particular plant virus species or strain (22, 23, 25, 28, 89–91). The vector species-specific interactions with plant viruses is determined by the recognition of protein binding sites in the virus that are recognized by insect receptors, which are thought to mediate the translocation of virions from the digestive tract to the hemolymph and salivary glands. Very little is known about aphid receptors of plant viruses. The first Luteovirid receptor in an aphid vector was identified as the aphid membrane alanyl aminopeptidase N, a cell surface receptor for *Pea enation mosaic virus* (PEMV, genus *Enamovirus*) in the pea aphid, *Acyrthosiphon pisum* (74, 75). Due to the biochemical nature of the interaction between aphids and plant viruses, proteomics approaches have been widely used to advance our knowledge on how the interplay works and what proteins are directly or indirectly involved in the interaction. These proteins are potential targets for management strategies aiming to block the ability of aphids to transmit plant viruses.

The search for potential cellular receptors to plant viruses in aphids has been going on for years. In 2001, Li and collaborators isolated two proteins from head tissues of *S. avenae* that were identified as potential receptors for BYDV-MAV based on virus overlay assays and 2-D (two-dimensional electrophoresis) immunoblot assays (92). However, the identity of the proteins was not investigated using mass spectrometry (MS). Vector biology studies have significantly benefitted from the use of proteomics approaches, especially if coupled with genetics approaches. Aphids are an excellent model organism for this type of study because they alternate sexual reproduction with parthenogenesis. Sexual reproduction is induced in aphids during the fall season, when sexual morphs are generated and aphids lay eggs for overwintering. Because aphids reproduce

parthenogenetically, they can be maintained as clonal lineages under laboratory conditions, so that an aphid colony started from a single female will produce only genetically identical individuals. Sexual reproduction can be induced in the lab reducing the daylight and temperature, obtaining male and female sexual morphs that can be used in crosses between two different aphid biotypes from the same species, generating an F1 population. This approach was used by Papura and collaborators to generate 39 F1 clones by selfing a poorly efficient BYDV-PAV clone of *S. avenae*, which were subsequently maintained as individual parthenogenetic colonies and used in transmission assays (91). The F1 clones segregated for the transmission of BYDV-PAV in a continuum from 0% to 88%. A 2-D gel approach was applied for proteome comparison of three of the most efficient vectors and two of the poorest vectors, obtaining four protein spots that were differentially expressed between the two groups, two of them with the same molecular weight but differing in their isoelectric points. Again, in this study, a MS approach was not used to identify the protein spots found to be involved in virus transmission by aphids. As the technology has become more accessible, the use of mass spectrometry (MS) to identify protein spots differentially expressed in 2-D assays has contributed for advancing our knowledge on the proteins involved in the virus transmission phenotype in aphids (27, 28, 93). For example, Seddas and collaborators used SDS-PAGE or 2-D gel to separate proteins from *M. persicae* that were then used in a far-western blot assay to test for binding to BWYV purified particles (93). Selected spots were then identified by MS as the aphid proteins Rack-1, GADPH3 and actin and hypothesized to be involved in the transcytosis of virus particles in the aphid (93).

In a series of elegant experiments, the Gray lab generated an F1 and F2

populations from a cross between two genotypes of *S. graminum*, one an efficient vector and the other a poor vector of BYDV-SGV and CYDV-RPV (22, 23). Using aphid genetics, the authors showed that the transmission of these two viruses by *S. graminum* is a genetically controlled and heritable trait and that the parents used in the cross are heterozygous for the genes controlling virus transmission. They also showed that a major gene and a few genes are involved in the virus transmission phenotype in an additive fashion and that tissue-specific barriers to virus transmission in the gut are not genetically linked to the barrier in the accessory salivary glands. Later, the F2 genotypes of *S. graminum* produced in these studies were used for the identification of proteins involved in virus transmission in *S. graminum*, by combining genetics analysis with 2D-DIGE (Fluorescence difference gel electrophoresis) and tandem MS for the identification of differentially expressed protein spots (24, 27, 28). This proteomics approach is one of the most widely used platforms for quantification of intact proteins. In one of these studies, four aphid proteins were associated with the ability of *S. graminum* to transmit CYDV-RPV and these proteins coimmunoprecipitated with purified CYDV-RPV (27). Two of these proteins were identified by MS as luciferase and cyclophilin, both involved in macromolecular transport. The other two proteins did not have any homologous proteins in the databases searched. The genome of *S. graminum* is not yet sequenced and at the time this study was performed, the only aphid database available was an EST database of the pea aphid *A. pisum* (the pea aphid genome was published in 2010). One of the limitations of proteomics approaches is the limited genomics resources for most insect species. For most aphid species, protein identification is performed based on homology to known proteins, mainly the *Drosophila* database, since the pea aphid genome is not

completely annotated yet. However, *Drosophila* differs from aphids in many aspects of their biology and this has implications in the level of similarity of proteins for homology-based searches. The draft genome of *M. persicae*, is being prepared for publication by the International consortium (aphidbase.org), but the annotation is still in the beginning. The genome of another important Hemipteran species, the whitefly *Bemisia tabaci*, has been recently published (94). The publication of the genome of these two Hemipteran species will significantly contribute to the power of proteomics searches for future studies.

After the publication of the pea aphid genome in 2010, the Gray lab performed another proteomics study comparing the F2 genotypes of *S. graminum* obtained by the cross between a vector and a nonvector of B/CYDV, for the identification of proteins involved in virus trasmission (28). Using a new customized database, the authors identified a total of 50 aphid's and *Buchnera*'s proteins that were differentially expressed between the two groups of efficient and poor vectors of B/CYDV. This was the first study to show that protein isoforms from the bacterial endosymbiont of aphids contribute to the virus transmission phenotype, in an additive manner. The aphid proteins identified to be involved in the virus transmission phenotype were related to innate immunity, protein folding and cell transport (28). In another study, a subset of these proteins was identified as protein biomarkers to predict whether a population of aphids or whiteflies would be efficient or poor at transmitting plant viruses (24). The identification of protein biomarkers is another promising utilization of proteomics approaches in insect management. The proteins biomarkers identified in this study included coA ligase, dihydropteridine reductase, cyclophilin, troponin-T, Rep70A, PITP, SERPIN-4, ETF-dehydrogenase, ROP-2 and a cuticle protein more recently annotated as a tyrosine-protein phosphatase

non-receptor type 23-like from *A. pisum* (24, 26, 28, 95). Using a 2D-DIGE coupled with mass spectrometry proteomics approach, aphid and *Buchnera* proteins involved in the virulence phenotype of *S. graminum* were identified. Proteomics approaches have been proved to be a useful tool for the identification of aphid proteins involved in virulence and plant virus transmission by aphids.

Bacterial endosymbionts of aphids

Symbiosis with bacteria is common in many insect species. In insects of the order Hemiptera, bacterial endosymbionts play an important role of synthesizing essential amino acids that are not obtained by the insects from their phloem sap diet. These endosymbionts are localized in specialized insect cells, as the bacteriocytes in aphids and whiteflies and the filter chamber that connects the whitefly midgut to hindgut, which is a site of passage for circulative viruses. Indeed, plant viruses that are transmitted by insects in a circulative manner interact with different insect tissues during the process of circulative transmission.

An area of intense debate in the field is whether bacterial symbionts of insect vectors are involved in the virus transmission process. Pinheiro and colleagues recently proposed a model to test for the involvement of the bacterial endosymbiont of insect vectors on their ability to transmit viruses (8). Although a direct involvement of *Buchnera* in the transmission of circulative viruses (such as Luteovirids) by aphids has been called into question (96, 97), evidence that a specific *Buchnera* genotype is related to the ability of the *S. graminum* to transmit CYDV-RPV suggests an indirect involvement of the

endosymbiont in virus transmission by aphids (28). Considering that the aphid-*Buchnera* association is obligate for both organisms, the involvement of *Buchnera* in the vectoring ability phenotype of aphids suggests a collaboration between aphids, bacteria and Luteovirids as a superorganism.

Are aphids, their endosymbionts and Luteovirids a superorganism?

Aphids and their bacterial endosymbiont *Buchnera* depend on each other for survival [3]. The relationship is an extreme example of cooperation between species that can now be probed in an unparalleled depth at the molecular level since the genomes of both organisms are sequenced [4-7]. It is therefore easy to formulate a hypothesis where the aphid-endosymbiont pair is an example of a superorganism along a continuum of "superorganismness". Interactions between aphids and host plants represent the opposite end of this continuum. On the other hand, a novel concept is that the superorganism idea can be applied to the aphid-Buchnera-Luteovirid system. Although aphids do not depend on vectored plant viruses for survival, it is known that aphids benefit from feeding on a plant infected with the circulative viruses they transmit, remarkably with Luteovirids.(11). Also, for an aphid species vector of a certain plant virus, it is rare to find non-vector populations in the field, suggesting that the vectoring phenotype has been favored by natural selection. Another interesting piece of information is that a particular genotype of Buchnera has been associated with the ability to efficiently transmit the Luteovirid *Cereal yellow dwarf virus* (CYDV), suggesting an indirect effect of Buchnera in virus transmission by aphids (28). On the other side of this interaction, Luteovirids depend

exclusively on their aphid vectors for transmission and it has been shown that these viruses influence the aphid feeding behavior to increase their plant-to-plant spread (fitness), in a very efficient system. Also, Luteovirids are transported across aphid tissues without causing any harm to aphids due to their lack of replication in aphid tissues. Altogether, these information suggests a mutualistic relationship between aphids and Luteovirids (8). It is also known that other sap sucking insects, closely related to aphids, such as the whitefly *Bemisia tabaci*, and the plant viruses they transmit in a circulative manner, such as Begomoviruses, collaborate in fighting host plant defenses (6, 98), which is likely to occur with the system aphids-Luteovirids as well.

A question that is guaranteed to stimulate fervent discussion among evolutionary biologists is: “Do superorganisms even exist?” The two possible definitive answers to this question often parallel views about the relative importance of “within-group l selection” and “between-group selection” in nature, thereby connecting to yet another raging debate within evolutionary biology [8]. We will answer the question first by making precise exactly what the question means in a simple game-theoretical framework. The term “superorganism” is usually reserved for cooperative societies composed of kin, but we show how it can be applied to non-kin, indeed different species, that have overlapping genetic interests because they reproduce in parallel.

The first task is to define precisely what is meant by a “superorganism”. The term usually conjures up an image of a group of organisms organized into a higher-level unit that develops and behaves much like a higher-level organism by exhibiting extensive internal cooperation manifested as an ergonomic division of labor (corresponding to somatic differentiation) and a division of reproduction (corresponding to the germline-

somatic cell distinction) among its component individuals. The latter picture appears close to what the social insect biologists Wheeler [24] and Emerson [25] meant when they visualized eusocial insect societies as superorganisms in which the altruistic workers were akin to somatic cells and the queen comparable to the gonads of a single organism. The latter is not the case when you have cooperating species, as with aphids and *Buchnera*, since logically neither party can relinquish reproduction. Although the concept of a superorganism first emerged from the social insect literature, it can readily be applied to co-reproducing species with an overlap in genetic interests caused by co-reproduction. Note that, under this conception, the biological universe is no longer dichotomized into "superorganisms" versus "non-superorganisms" but rather the "superorganismness" of a group is seen as being described by a point along a continuum ranging from the complete lack of to a perfect overlap in the genetic interests of the interacting component individuals. Notice that there is no requirement in this definition of superorganism that the component individuals be (i) aggregated in space or (ii) of the same species. Moreover, the evolutionary definition above does not necessarily require that there is be a substantial reproductive asymmetry or an elaborate task specialization among group members for a group to score high on the superorganismness continuum. All that is required is that the group members on average invest a large proportion of time or energy to promote group reproductive output (e.g., by enhancing the group's ability to capture resources) at the expense of their within-group share of resources.

How could a group of organisms possibly evolve to the upper extreme of the superorganism continuum? This is the empirically fruitful question that has occupied theoretical evolutionary biologists for several decades. Theory tells us that this might

come about through natural selection in several ways that are not mutually exclusive. All of the selective mechanisms described below have the common property that they cause increasing alignment in the genetic interests of the individuals making up the group, thereby driving the group upward along the superorganism continuum. More precisely, they favor the spread, through natural selection, of genes that have the phenotypic effect of increasing the proportion of an energy or time budget that is allocated to promoting group output as opposed to increasing one's share of that output. We let *S*, the degree of superorganismness, simply be the energy allocated to increasing group output at the expense of increasing one's share of that output. The factors promoting a high degree of superorganismness are:

Kinship: Positive genetic relatedness among interacting individuals can increase S through kin selection because the genes promoting intra-group cooperation are indirectly promoting copies of themselves by aiding relatives, which tend to possess copies of those same genes [3]. Indeed, all of the most spectacular examples of high-S insect societies such as honey bee, termite and ant societies involve groups of kin [4,5].

Ecological factors such as those promoting inter-group competition: Game theory models show that resource competition between groups can greatly increase *S* by favoring the suppression of any within-group selfishness that would compromise a group's ability to compete with other groups [6], and fierce inter-group competition for food and high-quality nest sites may lie behind the evolution of the highly elaborated patterns of cooperation and communication within ant societies [7]. (An unfortunately common notion is that the importance of inter-group competition in driving intra-group cooperation entails that "group selection" is stronger than "individual selection" in nature.

However, the latter is true only for “group selection” models in which groups are formed within breeding populations (i.e., groups are not different breeding populations). In such cases the value of *S* that evolves is the same as that which simultaneously maximizes the fitness of all individuals, so these models can equally well be seen as individual selection models [6,8]).

Co-reproduction. Even when group members are not genetically related or even of the same species, their reproductive interests can be parallel, as (i) when successful reproduction by one party requires successful reproduction by the other (as in plant-pollinator systems) or (ii) when the two parties reproduce together in the same way (as when fungi in the garden of a fungus-growing ant colonies is transmitted vertically to a new daughter colony via dispersal by reproductive ants from the parent colony). The high level of *S* exhibited by chloroplasts and mitochondria and their host eukaryotic cells (to the point that these are no longer seen as distinct organisms) no doubt reflects the fact that their genetic interests are entirely overlapping due to their parallel vertical transmission. This also is precisely the idea behind the application of the superorganism concept to the aphid-Luteovirid-*Buchnera* system.

Suppose in a host-parasite system, both the host and the parasite must each decide what fraction of their private resource (e.g., energy) pool to invest in generating a bigger, joint resource pool that will be used for reproduction. They also must each decide how much resource to invest selfishly in increasing their fraction of the reproductive resource pool against the selfish effort of the other party. Let *S* be the fraction of its private resource that the host invests in augmenting the group resource and *S*’ be that for the parasite. Thus, the group resource becomes $ST + S'T'$, if the host and parasite

start out with private energy stores *T* and *T'*, respectively. It follows that the host invests *(1-S)T* and the parasite invests *(1-S)T'* in the tug-of-war over their shares of the group resource. The host's reproduction *W* is assumed to be proportional to its fraction of the group resource won in the tug-of-war times the total amount of group resource generated:

$$W = \left(\frac{(1-S)T}{(1-S)T + (1-S')T'}\right)[ST + S'T'] \qquad (1)$$

The parasite's reproduction number of progeny will be its fraction of the group resource won in the tug-of-war times the total amount of group resource generated. However, the parasite's total fitness *W'* will be its total number of progeny times the success of transmitting those progeny to new hosts, and the latter can depend on the fitness of the host, if having a healthy, reproducing host is important for transmitting the parasite to new hosts. Thus, parasite's total fitness constructed as

$$W' = \left\{\left(\frac{(1-S')T'}{(1-S)T + (1-S')T'}\right)[ST + S'T']\right\} W^u \qquad (2)$$

where the probability of successful transmission of its progeny is assumed to have a possible dependence on the resources acquired by the host, the strength of the dependence being captured by the exponent *u*. If there is no such dependence, *u* = 0, and higher values of *u* mean that there is greater dependence of parasite transmission the greater is the host's amount of group resource won in the tug-of-war.

The evolutionarily stable investments in the group, *S** and *S'**, are those that jointly maximize both *W* and *W'* and are given by:

$$S^* = 1 - \frac{(T + T')(1 + u)}{T(2 + u)^2} \quad , \qquad S'^* = 1 - \frac{(T + T')}{T'(2 + u)^2} \qquad (3)$$

From the solutions in (3), increasing dependence of the parasite's transmission success on the resources of the host, *u*, increases the cooperative investments of both the host and the parasite in the group. In fact, for large *u*, both parasite and host invest virtually all of their private resource in cooperation, and an extreme "superorganismness" results.

The empirical results from the aphid proteomics literature indicate that the above superorganism model indeed can be applied to systems like the aphid-virus-*Buchnera* one. As noted earlier, cooperation in host-parasite systems seems to correlate positively with the parasite's dependence on a health, reproducing host for its own transmission. The latter condition certainly applies to the aphid-*Buchnera* pair. The sequencing of the aphid and *Buchnera* genomes has enabled researchers to paint an elegant picture of the high degree of cooperative genomic and proteomic interplay between the partners of the aphid superorganism as they relate to metabolic symbiosis [9-13], immune system co-adaptation [4, 14], and plant virus transmission [15, 16] and to starkly contrast this image with proteomic signatures of conflict between aphids and their host plants [17-23]. In this situation, where the aphid and Buchnera become one superorganism (the host), the two-way model can apply to the interaction between aphids and the luteovirids.

Vector manipulation by plant viruses versus collaboration

Pathogens and parasites interact with their hosts and vectors in many cellular and molecular pathways, which potentially causes changes in the host or vector behavior, in benefit of the pathogen spread and fitness (10). Recently, the vector manipulation

hypothesis has been proposed to explain the relationship between insect vectors and the plant viruses they transmit (10, 99), in parallel to what is known about animal viruses influencing their hosts' behavior (100–105). The "vector manipulation hypothesis" predicts that a virus will promote its plant-to-plant spread by influencing the plant host selection and feeding behavior of the insect vector (10, 99) to the detriment of the host and vector. Both persistent and nonpersistent viruses alter the host plant to make it more attractive to insect vectors, but they have different strategies to promote their plant-to-plant spread. Persistently transmitted viruses alter the host plant to enhance plant quality, so that the insect vector will have a long term feeding on the infected plant (106, 107), which will potentially make aphids acquire more virions and become more viruliferous. In contrast, it has been suggested that nonpersistently transmitted viruses influence the plant phenotype to make it a reduced quality diet, thereby promoting a rapid insect vector dispersal (108, 109). In both strategies, upon virus acquisition, insect vectors become more attracted to healthy plants, which completes the plant-to-plant spread cycle of the plant virus. However, the use of the term "manipulation" suggests a conflicting relationship, where plant viruses would alter the insect vector behavior for the virus benefit and in detriment of the insect vector. That does not seem to be the case for the aphids-Luteovirids system, which has been shown to benefit from their interaction, suggesting a mutualistic instead of a conflicting association.

Positive or neutral effects on vector performance have been extensively reported for persistently transmitted viruses that are dependent on their insect vectors for transmission (110–119). On the other hand, negative and sometimes neutral effects on insects have been reported mainly for insects feeding on plants infected with viruses and

other pathogens that are not transmitted or not exclusively transmitted by the insect species studied (4, 120–124). In the case of viruses that replicate in their insect vectors, such as the relationship between mosquitoes and animal viruses, detrimental effects have been shown for the vectors (125), as well as the manipulation of the vector by the virus (126), for the virus benefit. Collectively, these studies show that the extent to what viruses and insect vectors collaborate is related to the mode of transmission, as explained by the evolution of host virulence (127). Thus, more collaborative interactions will occur between insect vectors and viruses that depend exclusively on the vector for transmission, which might also be related to the length of time of their evolutionary association. In light of this superorganism paradigm, we propose that the term “vector manipulation” be replaced by “collaboration” in the aphid-Buchnera-Luteovirid system, as well as in other systems where the plant virus is transmitted in a circulative manner by the insect vector and with evidence of a mutualistic interaction, such as the whitefly-endosymbiont-Begomovirus system. This novel concept will help us improve our understanding of these complex interactions aiming to block the ability of aphids and plant viruses to interact, thus reducing virus spread in the field.

References

1. **Goggin FL**. 2007. Plant–aphid interactions: molecular and ecological perspectives. Curr Opin Plant Biol **10**:399–408.

2. **Goggin FL**, **Williamson VM**, **Ullman DE**. 2001. Variability in the Response of Macrosiphum euphorbiae and Myzus persicae (Hemiptera: Aphididae) to the Tomato Resistance Gene Mi. Environ Entomol **30**:101–106.

3. **Hebert SL**, **Jia L**, **Goggin FL**. 2007. Quantitative Differences in Aphid Virulence and Foliar Symptom Development on Tomato Plants Carrying the Mi Resistance Gene. Environ Entomol **36**:458–467.

4. **Hodge S**, **Powell G**. 2008. Do plant viruses facilitate their aphid vectors by inducing symptoms that alter behavior and performance? Environ Entomol **37**:1573–1581.

5. **Hodge S**, **Powell G**. 2010. Conditional facilitation of an aphid vector, Acyrthosiphon pisum, by the plant pathogen, pea enation mosaic virus. J Insect Sci **10**:155.

6. **Jiu M**, **Zhou X-P**, **Tong L**, **Xu J**, **Yang X**, **Wan F-H**, **Liu S-S**. 2007. Vector-virus mutualism accelerates population increase of an invasive whitefly. PLoS ONE **2**:e182.

7. **Kersch-Becker MF**, **Thaler JS**. 2014. Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores. Oecologia **174**:883–892.

8. **Pinheiro PV**, **Kliot A**, **Ghanim M**, **Cilia M**. 2015. Is there a role for symbiotic bacteria in plant virus transmission by insects? Curr Opin Insect Sci **8**:69–78.

9. **Castle SJ**, **Mowry TM**, **Berger PH**. 1998. Differential settling by Myzus persicae (Homoptera: Aphididae) on various virus infected host plants. Ann Entomol Soc Am **91**:661–667.

10. **Ingwell LL**, **Eigenbrode SD**, **Bosque-Pérez NA**. 2012. Plant viruses alter insect behavior to enhance their spread. Sci Rep **2**:578.

11. **Castle SJ**, **Berger PH**. 1993. Rates of growth and increase of Myzus persicae on virus-infected potatoes according to type of virus-vector relationship. Entomol Exp Appl **69**:51–60.

12. **Blackman R**, **Eastop V**. 2000. Wiley: Aphids on the world's crops: an identification and information guide, 2nd Edition - R. L. Blackman, V. F. Eastop.

13. **Miles PW**. 1999. Aphid saliva. Biol Rev **74**:41–85.

14. **Gray S**, **Cilia M**, **Ghanim M**. 2014. Circulative, "nonpropagative" virus transmission: an orchestra of virus-, insect-, and plant-derived instruments. Adv Virus Res **89**:141–199.

15. **Whitfield AE**, **Falk BW**, **Rotenberg D**. 2015. Insect vector-mediated transmission of plant viruses. Virology **479**–**480**:278–289.

16. **Nault LR**. 1997. Arthropod transmission of plant viruses: A new synthesis. Ann Entomol Soc Am **90**:521–541.

17. **Ng JCK**, **Perry KL**. 2004. Transmission of plant viruses by aphid vectors. Mol Plant Pathol **5**:505–511.

18. **Bak A**, **Martinière A**, **Blanc S**, **Drucker M**. 2013. Early interactions during the encounter of plants, aphids and arboviruses. Plant Signal Behav **8**.

19. **Pinheiro PV**, **Ghanim M**, **Alexander MMM**, **Rebelo AR**, **Santos RS**, **Orsburn BC**, **Gray S**, **Cilia M**. 2016. Host plants indirectly influence plant virus transmission by altering gut cysteine protease activity of aphid vectors. Mol Cell Proteomics mcp.M116.063495.

20. **Burd JD**, **Porter DR**. 2006. Biotypic diversity in greenbug (Hemiptera: Aphididae): characterizing new virulence and host associations. J Econ Entomol **99**:959–965.

21. **Porter DR**, **Burd JD**, **Shufran KA**, **Webster JA**. 2000. Efficacy of Pyramiding Greenbug (Homoptera: Aphididae) Resistance Genes in Wheat. J Econ Entomol **93**:1315–1318.

22. **Burrows ME**, **Caillaud MC**, **Smith DM**, **Benson EC**, **Gildow FE**, **Gray SM**. 2006. Genetic Regulation of Polerovirus and Luteovirus Transmission in the Aphid Schizaphis graminum. Phytopathology **96**:828–837.

23. **Burrows ME**, **Caillaud MC**, **Smith DM**, **Gray SM**. 2007. Biometrical genetic analysis of luteovirus transmission in the aphid Schizaphis graminum. Heredity **98**:106–113.

24. **Cilia M**, **Howe K**, **Fish T**, **Smith D**, **Mahoney J**, **Tamborindeguy C**, **Burd J**, **Thannhauser TW**, **Gray S**. 2011. Biomarker discovery from the top down: Protein biomarkers for efficient virus transmission by insects (Homoptera: Aphididae) discovered by coupling genetics and 2-D DIGE. Proteomics **11**:2440–2458.

25. **Gray SM**, **Caillaud MC**, **Burrows M**, **Smith DM**. 2007. Transmission of two viruses that cause Barley Yellow Dwarf is controlled by different loci in the aphid, Schizaphis graminum. J Insect Sci Online **7**:1–15.

26. **Tamborindeguy C**, **Bereman MS**, **DeBlasio S**, **Igwe D**, **Smith DM**, **White F**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2013. Genomic and proteomic analysis of Schizaphis graminum reveals cyclophilin proteins are involved in the transmission of cereal yellow dwarf virus. PloS One **8**:e71620.

27. **Yang X**, **Thannhauser TW**, **Burrows M**, **Cox-Foster D**, **Gildow FE**, **Gray SM**. 2008. Coupling genetics and proteomics to identify aphid proteins associated with vector-specific transmission of polerovirus (luteoviridae). J Virol **82**:291–299.

28. **Cilia M**, **Tamborindeguy C**, **Fish T**, **Howe K**, **Thannhauser TW**, **Gray S**. 2011. Genetics Coupled to Quantitative Intact Proteomics Links Heritable Aphid and Endosymbiont Protein Expression to Circulative Polerovirus Transmission. J Virol **85**:2148–2166.

29. **Gray SM**, **Smith DM**, **Barbierri L**, **Burd J**. 2002. Virus Transmission Phenotype Is Correlated with Host Adaptation Among Genetically Diverse Populations of the Aphid Schizaphis graminum. Phytopathology **92**:970–975.

30. **Pinheiro P**, **Bereman MS**, **Burd J**, **Pals M**, **Armstrong S**, **Howe KJ**, **Thannhauser TW**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2014. Evidence of the biochemical basis of host virulence in the greenbug aphid, Schizaphis graminum (Homoptera: Aphididae). J Proteome Res **13**:2094–2108.

31. **Shigenobu S**, **Wilson ACC**. 2011. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. Cell Mol Life Sci CMLS **68**:1297–1309.

32. **Douglas AE**. 1998. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. Annu Rev Entomol **43**:17–37.

33. **Wilkinson TL**, **Douglas AE**. 1996. The impact of aposymbiosis on amino acid metabolism of pea aphids, p. 279–282. *In* Städler, E, Rowell-Rahier, M, Bauer, R (eds.), Proceedings of the 9th International Symposium on Insect-Plant Relationships. Springer Netherlands.

34. **Poliakov A**, **Russell CW**, **Ponnala L**, **Hoops HJ**, **Sun Q**, **Douglas AE**, **van Wijk KJ**. 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. Mol Cell Proteomics MCP **10**:M110.007039.

35. **Buchner P**. 1965. Endosymbiosis of animals with plant microorganisms. Interscience Publishers.

36. **Miura T**, **Braendle C**, **Shingleton A**, **Sisk G**, **Kambhampati S**, **Stern DL**. 2003. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid Acyrthosiphon pisum (Hemiptera: Aphidoidea). J Exp Zoolog B Mol Dev Evol **295B**:59–81.

37. **Serbus LR**, **White PM**, **Silva JP**, **Rabe A**, **Teixeira L**, **Albertson R**, **Sullivan W**. 2015. The Impact of Host Diet on Wolbachia Titer in Drosophila. PLOS Pathog **11**:e1004777.

38. **International Aphid Genomics Consortium**. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol **8**:e1000313.

39. **Gerardo NM**, **Altincicek B**, **Anselme C**, **Atamian H**, **Barribeau SM**, **de Vos M**, **Duncan EJ**, **Evans JD**, **Gabaldón T**, **Ghanim M**, **Heddi A**, **Kaloshian I**, **Latorre A**, **Moya A**, **Nakabachi A**, **Parker BJ**, **Pérez-Brocal V**, **Pignatelli M**, **Rahbé Y**, **Ramsey JS**, **Spragg CJ**, **Tamames J**, **Tamarit D**, **Tamborindeguy C**, **Vincent-Monegat C**, **Vilcinskas A**. 2010. Immunity and other defenses in pea aphids, Acyrthosiphon pisum. Genome Biol **11**:R21.

40. **Douglas AE**, **Bouvaine S**, **Russell RR**. 2011. How the insect immune system interacts with an obligate symbiotic bacterium. Proc R Soc B Biol Sci **278**:333–338.

41. **Shigenobu S**, **Watanabe H**, **Hattori M**, **Sakaki Y**, **Ishikawa H**. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature **407**:81–86.

42. **Hansen AK**, **Moran NA**. 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. Proc Natl Acad Sci **108**:2849–2854.

43. **Hansen AK**, **Degnan PH**. 2014. Widespread expression of conserved small RNAs in small symbiont genomes. ISME J **8**:2490–2502.

44. **Powell G**, **Tosh CR**, **Hardie J**. 2006. HOST PLANT SELECTION BY APHIDS: Behavioral, Evolutionary, and Applied Perspectives. Annu Rev Entomol **51**:309–330.

45. **Tjallingii W**. 1994. Sieve element acceptance by aphids. EJE **91**:47–52.

46. **Peccoud J**, **Ollivier A**, **Plantegenest M**, **Simon J-C**. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. Proc Natl Acad Sci **106**:7495–7500.

47. **Gao S**, **Liu D**. 2013. Differential performance of Sitobion avenae (Hemiptera: Aphididae) clones from wheat and barley with implications for its management through alternative cultural practices. J Econ Entomol **106**:1294–1301.

48. **Lu H**, **Yang P**, **Xu Y**, **Luo L**, **Zhu J**, **Cui N**, **Kang L**, **Cui F**. 2016. Performances of survival, feeding behavior, and gene expression in aphids reveal their different fitness to host alteration. Sci Rep **6**:19344.

49. **García-Robledo C**, **Horvitz CC**. 2012. Jack of all trades masters novel host plants: positive genetic correlations in specialist and generalist insect herbivores expanding their diets to novel hosts. J Evol Biol **25**:38–53.

50. **Carolan JC**, **Caragea D**, **Reardon KT**, **Mutti NS**, **Dittmer N**, **Pappan K**, **Cui F**, **Castaneto M**, **Poulain J**, **Dossat C**, **Tagu D**, **Reese JC**, **Reeck GR**, **Wilkinson TL**, **Edwards OR**. 2011. Predicted effector molecules in the salivary secretome of the pea aphid (Acyrthosiphon pisum): a dual transcriptomic/proteomic approach. J Proteome Res **10**:1505–1518.

51. **Hogenhout SA**, **Bos JIB**. 2011. Effector proteins that modulate plant--insect interactions. Curr Opin Plant Biol **14**:422–428.

52. **Weber G**. 1985. Genetic variability in host plant adaptation of the green peach aphid, Myzus persicae. Entomol Exp Appl **38**:49–56.

53. **Takada H**. 1981. Inheritance of Body Colors in Myzus persicae (SULZER)(Homoptera : Aphididae). Appl Entomol Zool **16**:242–246.

54. **Blackman RL**. 1974. Life-cycle variation of Myzus persicae (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. Bull Entomol Res **63**:595–607.

55. **Margaritopoulos JT**, **Tsitsipis JA**, **Goudoudaki S**, **Blackman RL**. 2002. Life cycle variation of Myzus persicae (Hemiptera : Aphididae) in Greece. Bull Entomol Res **92**:309–319.

56. **Blackman RL**, **Takada H**. 1975. A naturally occurring chromosomal translocation in Myzus persicae (Sulzer). J Entomol A **50**:147–156.

57. **Kempton RA**, **Lowe HJB**, **Bintcliffe EJB**. 1980. The Relationship between Fecundity and Adult Weight in Myzus persicae. J Anim Ecol **49**:917–926.

58. **Lowe H**. 1974. Intraspecific Variation of Myzus-Persicae on Sugar-Beet (beta-Vulgaris). Ann Appl Biol **78**:15–26.

59. **Francis F**, **Gerkens P**, **Harmel N**, **Mazzucchelli G**, **De Pauw E**, **Haubruge E**. 2006. Proteomics in Myzus persicae: effect of aphid host plant switch. Insect Biochem Mol Biol **36**:219–227.

60. **Symmes EJ**, **Perring TM**. 2007. Intraspecific variation in Zucchini yellow mosaic virus transmission by Myzus persicae and the impact of aphid host plant. J Econ Entomol **100**:1764–1772.

61. **Lucio-Zavaleta E**, **Smith DM**, **Gray SM**. 2001. Variation in Transmission Efficiency Among Barley yellow dwarf virus-RMV Isolates and Clones of the Normally Inefficient Aphid Vector, Rhopalosiphum padi. Phytopathology **91**:792–796.

62. **Gladders DW**, **Peters D**. 1986. The effect of previous host plant on the fecundity of Myzus persicae and its ability to transmit beet yellows virus. Ann Appl Biol **109**:499–507.

63. **Mac Kinnon JP**. 1963. Some factors that affect the aphid transmission of two viruses that persist in the vector. Virology **20**:281–287.

64. **Tamada T**, **Harrison BD**. 1981. Quantitative studies on the uptake and retention of Potato leafroll virus by aphids in laboratory and field conditions. Ann Appl Biol **98**:261–276.

65. **Cilia M**, **Peter KA**, **Bereman MS**, **Howe K**, **Fish T**, **Smith D**, **Gildow F**, **MacCoss MJ**, **Thannhauser TW**, **Gray SM**. 2012. Discovery and targeted LC-MS/MS of purified polerovirus reveals differences in the virus-host interactome associated with altered aphid transmission. PloS One **7**:e48177.

66. **Brault V**, **Périgon S**, **Reinbold C**, **Erdinger M**, **Scheidecker D**, **Herrbach E**, **Richards K**, **Ziegler-Graff V**. 2005. The polerovirus minor capsid protein determines vector specificity and intestinal tropism in the aphid. J Virol **79**:9685–9693.

67. **Reinbold C**, **Herrbach E**, **Brault V**. 2003. Posterior midgut and hindgut are both sites of acquisition of Cucurbit aphid-borne yellows virus in Myzus persicae and Aphis gossypii. J Gen Virol **84**:3473–3484.

68. **Gildow FE**, **Reavy B**, **Mayo MA**, **Duncan GH**, **Woodford JA**, **Lamb JW**, **Hay RT**. 2000. Aphid acquisition and cellular transport of Potato leafroll virus-like particles lacking P5 Readthrough protein. Phytopathology **90**:1153–1161.

69. **Garret A**, **Kerlan C**, **Thomas D**. 1996. Ultrastructural study of acquisition and retention of potato leafroll luteovirus in the alimentary canal of its aphid vector, Myzus persicae Sulz. Arch Virol **141**:1279–1292.

70. **Garret A**, **Kerlan C**, **Thomas D**. 1991. Visualization of a luteovirus in the vector aphid's body by two gold immunolabelling techniques: a comparative study. J Virol Methods **35**:93–107.

71. **Rouzé-Jouan J**, **Terradot L**, **Pasquer F**, **Tanguy S**, **Giblot Ducray-Bourdin D D**. 2001. The passage of Potato leafroll virus through Myzus persicae gut membrane regulates transmission efficiency. J Gen Virol **82**:17–23.

72. **Garret A**, **Kerlan C**, **Thomas D**. 1993. The intestine is a site of passage for potato leafroll virus from the gut lumen into the haemocoel in the aphid vector, Myzus persicae Sulz. Arch Virol **131**:377–392.

73. **Gildow FE**. 1987. Virus—membrane interactions involved in circulative transmission of luteoviruses by aphids, p. 93–120. *In* Harris, KF (ed.), Current Topics in Vector Research. Springer New York.

74. **Linz LB**, **Liu S**, **Chougule NP**, **Bonning BC**. 2015. In vitro evidence supports membrane alanyl aminopeptidase n as a receptor for a plant virus in the pea aphid vector. J Virol JVI.01479-15.

75. **Liu S**, **Sivakumar S**, **Sparks WO**, **Miller WA**, **Bonning BC**. 2010. A peptide that binds the pea aphid gut impedes entry of Pea enation mosaic virus into the aphid hemocoel. Virology **401**:107–116.

76. **Shepardson S**, **Esau K**, **McCrum R**. 1980. Ultrastructure of potato leaf phloem infected with potato leafroll virus. Virology **105**:379–392.

77. **Peiffer ML**, **Gildow FE**, **Gray SM**. 1997. Two distinct mechanisms regulate luteovirus transmission efficiency and specificity at the aphid salivary gland. J Gen Virol **78 (Pt 3)**:495–503.

78. **Gildow FE**. 1993. Evidence for receptor-mediated endocytosis regulating luteovirus acquisition by aphids. Phytopathol USA.

79. **Miller WA**, **Dinesh-Kumar SP**, **Paul CP**. 1995. Luteovirus Gene Expression. Crit Rev Plant Sci **14**:179–211.

80. **Brault V**, **van den Heuvel JF**, **Verbeek M**, **Ziegler-Graff V**, **Reutenauer A**, **Herrbach E**, **Garaud JC**, **Guilley H**, **Richards K**, **Jonard G**. 1995. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. EMBO J **14**:650–659.

81. **Cheng S-L**, **Domier LL**, **D'Arcy CJ**. 1994. Detection of the Readthrough Protein of Barley Yellow Dwarf Virus. Virology **202**:1003–1006.

82. **Filichkin SA**, **Lister RM**, **McGrath PF**, **Young MJ**. 1994. In Vivo Expression and Mutational Analysis of the Barley Yellow Dwarf Virus Readthrough Gene. Virology **205**:290–299.

83. **Wang JY**, **Chay C**, **Gildow FE**, **Gray SM**. 1995. Readthrough Protein Associated with Virions of Barley Yellow Dwarf Luteovirus and Its Potential Role in Regulating the Efficiency of Aphid Transmission. Virology **206**:954–962.

84. **Reutenauer A**, **Ziegler-Graff V**, **Lot H**, **Scheidecker D**, **Guilley H**, **Richards K**, **Jonard G**. 1993. Identification of Beet Western Yellows Luteovirus Genes Implicated in Viral Replication and Particle Morphogenesis. Virology **195**:692–699.

85. **Bruyère A**, **Brault V**, **Ziegler-Graff V**, **Simonis M-T**, **Van den Heuvel JFJM**, **Richards K**, **Guilley H**, **Jonard G**, **Herrbach E**. 1997. Effects of Mutations in the Beet Western Yellows Virus Readthrough Protein on Its Expression and Packaging and on Virus Accumulation, Symptoms, and Aphid Transmission. Virology **230**:323–334.

86. **CHAY CA**, **GUNASINGE UB**, **DINESH-KUMAR SP**, **MILLER WA**, **GRAY SM**. 1996. Aphid Transmission and Systemic Plant Infection Determinants of Barley Yellow Dwarf Luteovirus-PAV are Contained in the Coat Protein Readthrough Domain and 17-kDa Protein, Respectively. Virology **219**:57–65.

87. **Reinbold C**, **Gildow FE**, **Herrbach E**, **Ziegler-Graff V**, **Gonçalves MC**, **van den Heuvel JFJM**, **Brault V**. 2001. Studies on the role of the minor capsid protein in

transport of Beet western yellows virus through Myzus persicae. J Gen Virol **82**:1995–2007.

88. **Peter KA**, **Gildow F**, **Palukaitis P**, **Gray SM**. 2009. The C Terminus of the Polerovirus P5 Readthrough Domain Limits Virus Infection to the Phloem. J Virol **83**:5419–5429.

89. **Gray SM**, **Banerjee N**. 1999. Mechanisms of Arthropod Transmission of Plant and Animal Viruses. Microbiol Mol Biol Rev **63**:128–148.

90. **Guo J-Q**, **Moreau J-P**, **Lapierre H**. 1996. VARIABILITY AMONG APHID CLONES OF RHOPALOSIPHUM PADI L. AND SITOBION AVENAE FABR. (HOMOPTERA: APHIDIDAE) IN TRANSMISSION OF THREE PAV ISOLATES OF BARLEY YELLOW DWARF VIRUSES. Can Entomol **128**:209–217.

91. **Papura D**, **Jacquot E**, **Dedryver CA**, **Luche S**, **Riault G**, **Bossis M**, **Rabilloud T**. 2002. Two-dimensional electrophoresis of proteins discriminates aphid clones of Sitobion avenae differing in BYDV-PAV transmission. Arch Virol **147**:1881–1898.

92. **Li C**, **Cox-Foster D**, **Gray SM**, **Gildow F**. 2001. Vector specificity of barley yellow dwarf virus (BYDV) transmission: identification of potential cellular receptors binding BYDV-MAV in the aphid, Sitobion avenae. Virology **286**:125–133.

93. **Seddas P**, **Boissinot S**, **Strub J-M**, **Van Dorsselaer A**, **Van Regenmortel MHV**, **Pattus F**. 2004. Rack-1, GAPDH3, and actin: proteins of Myzus persicae potentially involved in the transcytosis of beet western yellows virus particles in the aphid. Virology **325**:399–412.

94. **Chen W**, **Hasegawa DK**, **Kaur N**, **Kliot A**, **Pinheiro PV**, **Luan J**, **Stensmyr MC**, **Zheng Y**, **Liu W**, **Sun H**, **Xu Y**, **Luo Y**, **Kruse A**, **Yang X**, **Kontsedalov S**, **Lebedev G**, **Fisher TW**, **Nelson DR**, **Hunter WB**, **Brown JK**, **Jander G**, **Cilia M**, **Douglas AE**, **Ghanim M**, **Simmons AM**, **Wintermantel WM**, **Ling K-S**, **Fei Z**. 2016. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol **14**:110.

95. **Cilia M**, **Bereman M**, **Fish T**, **MacCoss MJ**, **Gray S**. 2012. Homopteran Vector Biomarkers for Efficient Circulative Plant Virus Transmission are Conserved in Multiple Aphid Species and the Whitefly Bemisia tabaci. J Integr Agric **11**:249–262.

96. **Bouvaine S**, **Boonham N**, **Douglas AE**. 2011. Interactions between a luteovirus and the GroEL chaperonin protein of the symbiotic bacterium Buchnera aphidicola of aphids. J Gen Virol **92**:1467–1474.

97. **Liu S**, **Sivakumar S**, **Wang Z**, **Bonning BC**, **Miller WA**. 2009. The readthrough domain of pea enation mosaic virus coat protein is not essential for virus stability in the hemolymph of the pea aphid. Arch Virol **154**:469–479.

98. **Zhang T**, **Luan J-B**, **Qi J-F**, **Huang C-J**, **Li M**, **Zhou X-P**, **Liu S-S**. 2012. Begomovirus–whitefly mutualism is achieved through repression of plant defences by a virus pathogenicity factor. Mol Ecol **21**:1294–1304.

99. **Roosien BK**, **Gomulkiewicz R**, **Ingwell LL**, **Bosque-Pérez NA**, **Rajabaskar D**, **Eigenbrode SD**. 2013. Conditional vector preference aids the spread of plant pathogens: results from a model. Environ Entomol **42**:1299–1308.

100. **Hurd H**. 2003. Manipulation of Medically Important Insect Vectors by Their Parasites. Annu Rev Entomol **48**:141–161.

101. **Koella JC**, **Sørensen FL**, **Anderson RA**. 1998. The malaria parasite, Plasmodium falciparum, increases the frequency of multiple feeding of its mosquito vector, Anopheles gambiae. Proc R Soc Lond B Biol Sci **265**:763–768.

102. **Lacroix R**, **Mukabana WR**, **Gouagna LC**, **Koella JC**. 2005. Malaria Infection Increases Attractiveness of Humans to Mosquitoes. PLOS Biol **3**:e298.

103. **Poulin R**. 1995. “Adaptive” changes in the behaviour of parasitized animals: A critical review. Int J Parasitol **25**:1371–1383.

104. **Poulin R**. 2000. Manipulation of host behaviour by parasites: a weakening paradigm? Proc R Soc Lond B Biol Sci **267**:787–792.

105. **Thomas F**, **Adamo S**, **Moore J**. 2005. Parasitic manipulation: where are we and where should we go? Behav Processes **68**:185–199.

106. **Alvarez AE**, **Garzo E**, **Verbeek M**, **Vosman B**, **Dicke M**, **Tjallingii WF**. 2007. Infection of potato plants with potato leafroll virus changes attraction and feeding behaviour of Myzus persicae. Entomol Exp Appl **125**:135–144.

107. **Eigenbrode SD**, **Ding H**, **Shiel P**, **Berger PH**. 2002. Volatiles from potato plants infected with potato leafroll virus attract and arrest the virus vector, Myzus persicae (Homoptera: Aphididae). Proc Biol Sci **269**:455–460.

108. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. Proc Natl Acad Sci U S A **107**:3600–3605.

109. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Effects of Cucumber mosaic virus infection on vector and non-vector herbivores of squash. Commun Integr Biol **3**:579–582.

110. **Belliure B**, **Janssen A**, **Maris PC**, **Peters D**, **Sabelis MW**. 2005. Herbivore arthropods benefit from vectoring plant viruses. Ecol Lett **8**:70–79.

111. **Blua MJ**, **Perring TM**, **Madore MA**. 1994. Plant virus-induced changes in aphid population development and temporal fluctuations in plant nutrients. J Chem Ecol **20**:691–707.

112. **Castle SJ**, **Berger PH**. 1993. Rates of growth and increase of Myzus persicae on virus-infected potatoes according to type of virus-vector relationship. Entomol Exp Appl **69**:51–60.

113. **Luan J-B**, **Li J-M**, **Varela N**, **Wang Y-L**, **Li F-F**, **Bao Y-Y**, **Zhang C-X**, **Liu S-S**, **Wang X-W**. 2011. Global analysis of the transcriptional response of whitefly to tomato yellow leaf curl China virus reveals the relationship of coevolved adaptations. J Virol **85**:3330–3340.

114. **Luan J-B**, **Yao D-M**, **Zhang T**, **Walling LL**, **Yang M**, **Wang Y-J**, **Liu S-S**. 2013. Suppression of terpenoid synthesis in plants by a virus promotes its mutualism with vectors. Ecol Lett **16**:390–398.

115. **MacKinnon JP**. 1961. Preference of Aphids for Excised Leaves to Whole Plants. Can J Zool **39**:445–447.

116. **Maris PC**, **Joosten NN**, **Goldbach RW**, **Peters D**. 2004. Tomato spotted wilt virus infection improves host suitability for its vector Frankliniella occidentalis. Phytopathology **94**:706–711.

117. **Matsuura S**, **Hoshino S**. 2009. Effect of tomato yellow leaf curl disease on reproduction of Bemisia tabaci Q biotype (Hemiptera: Aleyrodidae) on tomato [Lycopersicon esculentum] plants. Appl Entomol Zool Jpn.

118. **Mayer RT**, **Inbar M**, **McKenzie CL**, **Shatters R**, **Borowicz V**, **Albrecht U**, **Powell CA**, **Doostdar H**. 2002. Multitrophic interactions of the silverleaf whitefly, host plants, competing herbivores, and phytopathogens. Arch Insect Biochem Physiol **51**:151–169.

119. **Miller JW**, **Coon BF**. 1964. The effect of Barley yellow dwarf virus on the biology of its vector the english grain aphid, Macrosiphum granarium. J Econ Entomol **57**:970–974.

120. **Nanayakkara UN**, **Giguère MA**, **Pelletier Y**. 2012. Population growth of Myzus persicae on potato plants infected with different strains and variants of potato virus Y. Am J Potato Res **90**:297–300.

121. **Wosula EN**, **Davis JA**, **Clark CA**. 2013. Population dynamics of three aphid species (Hemiptera: Aphididae) on four Ipomoea spp. infected or noninfected with sweetpotato potyviruses. J Econ Entomol **106**:1566–1573.

122. **Ellsbury MM**, **Pratt RG**, **Knight WE**. 1985. Effects of single and combined infection of arrowleaf clover with Bean yellow mosaic virus and a phytophthora sp. on reproduction and colonization by pea aphids (Homoptera: Aphididae). Environ Entomol **14**:356–359.

123. **Inoue T**, **Sakurai T**. 2006. Infection of *Tomato spotted wilt virus* (TSWV) shortens the life span of thelytokous *Thrips tabaci* (Thysanoptera: Thripidae). Appl Entomol Zool **41**:239–246.

124. **Donaldson JR**, **Gratton C**. 2007. Antagonistic effects of soybean viruses on soybean aphid performance. Environ Entomol **36**:918–925.

125. **Sim S**, **Ramirez JL**, **Dimopoulos G**. 2012. Dengue virus infection of the Aedes aegypti salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. PLoS Pathog **8**:e1002631.

126. **Mehlhorn H**. 2015. Host Manipulations by Parasites and Viruses. Springer.

127. **Ewald PW**. 1998. The evolution of virulence and emerging diseases. J Urban Health Bull N Y Acad Med **75**:480–491.

Chapter 1

Evidence for the biochemical basis of host virulence in the greenbug aphid, *Schizaphis graminum* (Homoptera: Aphididae)

Published as a research article in Journal of Proteome Research (2014) 13: 2094-2108

Patricia V. Pinheiro[1,2], Michael S. Bereman[3,§], John Burd[4], Melissa Pals[4], Scott Armstrong[4], Kevin J. Howe[5], Theodore W. Thannhauser[5], Michael J. MacCoss[3], Stewart M. Gray[5,6], Michelle Cilia[5,6,7,*]

[1] Department of Entomology, Cornell University, Ithaca, NY 14853 USA
[2] Embrapa Rice and Beans, Santo Antonio de Goias, GO, Brazil, 75375-000
[3] Department of Genome Sciences, University of Washington, Seattle, WA 98195 USA
[4] Wheat, Peanut and Other Field Crops Research Unit, USDA-ARS, 1301 N. Western Road, Stillwater, Ok 74075 USA
[5] Robert W. Holley Center for Agriculture and Health, USDA-ARS, Tower Road, Ithaca, NY 14853 USA
[6] Department of Plant Pathology and Plant Microbe Interactions, Cornell University, Ithaca, NY 14853 USA
[7] Boyce Thompson Institute for Plant Research, Ithaca, NY 14853 USA
§ Current affiliation: Department of Biological Sciences, Center for Human Health and the Environment, North Carolina State University, Raleigh, NC 27695 USA

*Correspondence should be addressed to: mlc68@cornell.edu

ABSTRACT

Biotypes of aphids and many other insect pests are defined based on the phenotypic response of host plants to the insect pest without considering their intrinsic characteristics and genotypes. Plant breeders have spent considerable effort to develop aphid-resistant, small grain varieties to limit insecticide control of the greenbug, *Schizaphis graminum*. However, new *S. graminum* biotypes frequently emerge that break resistance. Mechanisms of virulence on the aphid side of the plant-insect interaction are not well understood. *S. graminum* biotype H is highly virulent on most small grain varieties. This characteristic makes biotype H ideal for comparative proteomics to investigate the basis of biotype virulence in aphids. In this study, we used comparative proteomics to identify protein expression differences associated with virulence. Aphid proteins involved in the tricarboxylic acid cycle, immune system, cell division, and anti-apoptosis pathways were found to be up-regulated in biotype H relative to other biotypes. Proteins from the bacterial endosymbiont of aphids were also differentially expressed in biotype H. Guided by the proteome results, we tested whether biotype H had a fitness advantage compared to other *S. graminum* biotypes and found that biotype H had a higher reproductive fitness as compared to two other biotypes on a range of different wheat germplasms. Finally, we tested whether aphid genetics can be used to further dissect the genetic mechanisms of biotype virulence in aphids. The genetic data showed that sexual reproduction is a source of biotypic variation observed in *S. graminum.*

KEYWORDS: Aphid, biotype, vector biology, insect-plant interactions, host virulence, *Schizaphis graminum*, greenbug, DIGE, LC-MS/MS.

INTRODUCTION

Since 1931, the methodology to describe and categorize newly isolated

populations of many agricultural insect pests, including aphids, has remained the same for the most part (1). Reginald Painter described a biotype as a morphologically indistinct **"**subspecific strain**"** of an insect species that displayed a unique relationship with a genetically stable, resistant host plant (2). This method of insect biotype characterization relies on determining the phenotypic response of host plants to the insect pest without consideration for the insect genetic background. However, evolutionary theory predicts that co-evolutionary races between host and pests will typically yield genetic variability in *both* the plant defense strategies and counter-defense strategies used by the insect.

Aphids are Homopteran insects with piercing sucking mouthparts that feed directly from the plant vascular tissues (phloem) on the sap. Aphids have two unique advantages over other insect pests for studying co-evolutionary races between host plants and insects using proteomics approaches. First, aphids alternate sexual reproduction with parthenogenetic reproduction. Each parthenogenetic genotype represents a clone. Aphid hybrid lineages generated by sexual reproduction can be maintained parthenogenetically and allow us to phenotype each hybrid genotype using proteomics or other methods. Second, the genome sequence of the pea aphid, *Acyrthosiphon pisum*, has been published (3). Our studies (4–10) and others (11–13) highlight enormous benefits of the *A. pisum* genome for aphid functional genomics and proteomics. Recent intensive efforts to use proteomics approaches to study aphid saliva have revealed species-specific differences in salivary components that may emerge to be a major determinant of specificity in aphid-plant interactions (11, 12, 14–16). Aphid saliva contains proteins that can digest cell wall components and sugars, and that function in detoxification. Aphid salivary proteins may function as the first line of defense against the host plant immune

system and may even be considered as an arsenal of weapons that the aphid uses to help colonize host plants.

In the greenbug, *Schizaphis graminum*, the host plants used to determine biotype include varieties of the small grains, (wheat, barley, and rye), that have different resistance genes (17, 18). Phylogenetic analysis, and even mitochondrial DNA sequencing of different *S. graminum* biotypes reveals that the biotypes, as they are described by their host-plant response, are not always genetically discrete populations (19, 20). Thus, while the current system is widely used by and valuable to applied entomologists and plant breeders alike, this classification system is not grounded in evolutionary biology or insect genetics and is limited in scope by the host plants used in the assay. Furthermore, it falls short of distinguishing precisely between highly virulent pest biotypes. For example, as new *S. graminum* populations emerge and are classified, host plants previously thought to be resistant to multiple biotypes now show susceptibility to multiple different *S. graminum* populations (18, 19).

Greenbugs do more damage to cereal crops than caused by feeding alone. They are a major vector for plant viruses in the family Luteoviridae. Viruses in the Luteoviridae include the Barley and Cereal yellow dwarf viruses (B/CYDV), *Potato leafroll virus* (PLRV), *Beet western yellows virus* (BWYV) and *Soybean dwarf virus* among others. Although small in number, viruses in the Luteoviridae cause major economic losses to staple crops worldwide. These viruses are transmitted exclusively by aphids in a circulative-non-propagative manner, i.e., the viruses must circulate throughout the aphid body prior to transmission to a new host plant but they do not replicate in the insect nor are they passed on to progeny (21, 22). *S. graminum* transmits multiple species of

B/CYDV to a wide range of cereal crop hosts and native grasses (23–25). Natural selection has also favored B/CYDV to manipulate aphid behavior in a way that enhances virus spread (26, 27). B/CYDV transmission efficiency is a genetically controlled trait in *S. graminum* (28–30). Not all populations of *S. graminum* are efficient vectors for B/CYDVs (5, 6, 9, 10, 28–30).

Early work suggested that virus vectoring ability in *S. graminum* was correlated with host adaptation (31). *S. graminum* biotypes that were isolated from wild grasses were shown to be efficient vectors for different B/CYDV species, whereas those adapted to agronomic crops were weak vectors for B/CYDVs. The most efficient vector, biotype H, stood apart as it transmitted five species of yellow dwarf viruses (YDVs) with high efficiency (31) and is also the most virulent on agronomic host plants (17) independent of its ability to transmit virus. In host virulence assays, all but one wheat germplasm (Largo) succumbed to infestation by biotype H (17). These characteristics make biotype H an ideal biotype for comparative proteomics to probe the biochemical basis of biotype virulence in aphids. In this study, we used gel and liquid chromatography-based approaches to quantify proteome changes specifically associated with the virulence phenotype in biotype H as compared to 11 other *S. graminum* biotypes. This comprehensive proteome analysis implicates the involvement of the tricarboxylic acid (TCA) cycle, aphid immune system, anti-apoptosis signaling proteins as well as five proteins from the proteome of the primary bacterial endosymbiont, *Buchnera aphidicola*, in aphid virulence strategies. Proteomic variation in at least one of the bacterial proteins, Elongation Factor-Tu (EF-Tu), was explained by several single amino acid polymorphisms that conferred a change in the protein's isoelectric point. To test whether

the unique proteomic variation observed in biotype H was associated with a selective advantage compared to the other biotypes, we compared reproductive fitness of biotype H to two other biotypes on six different aphid-resistant wheat lines and one variety of barley. Biotype H had a larger number of progeny on all plants tested. The proteomic data indicate that biotype differences in *S. graminum* are the result of changes in the expression of multiple, yet a relatively small number of genes and proteins. To provide genetic support for the proteomic variation, we analyzed the biotype phenotypes of eight F2 genotypes as compared to the parental genotypes. The genetic data show that sexual reproduction is a rich source of biotypic diversity and confirm the proteomic observations that biotype determination is multigenic in *S. graminum*.

EXPERIMENTAL SECTION

Aphids

Parthenogenetically reproducing aphid colonies were maintained on caged barley (*Hordeum vulgare*) at 20°C with an 18-h photo period as described previously (31). The origin and CYDV transmission efficiency of the parental genotypes of Sg-SC and Sg-F, as well as the F2 genotypes A3, C2, K2 and K3 were described previously (6, 10, 28, 29, 32). All field-collected biotypes were started as first instar nymphs from adults reared in dishes to remove viruses. Nymphs were transferred to and reared on enclosed barley plants in walk-in growth chambers. The mating procedures to generate the F2 genotypes are described in refs (28, 29).

Virulence Assays

Wheat varieties Custer, DS 28A (Gb1), Amigo (Gb2), CI 17882 (Gb5), CI 17959 (Gb4), Largo (Gb3), GRS 1201 (Gb6), rye varieties Elbon, Insave (Gb2, Gb6), and barley varieties Wintermalt, Post 90 (Rsg1a) and PI 426756 (Rsg2b) seeds were planted in black plastic cone flats in Redi Earth using three seeds per cone, per variety. Four biological replicates were planted in each flat in a randomized design (Fig. 1.1A). When seedlings were approximately 2" tall, the flats were infested with the F2 genotype clone or biotype by placing infested plant leaves between rows of flat so that aphids were allowed to move over to desired seedlings. Eight to twelve days later, the plants were rated either live/dead to give rise to the resistant (live, R) or susceptible (dead, S) designation (Fig. 1.1B). A total of eleven *S. graminum* genotypes, including eight F2 genotypes, two parental genotypes and biotype H were tested in this assay (Table 1.1).

Figure 1.1 Plant reactions to *S. graminum* genotypes (R = resistant, S = susceptible). **A** Plants in a randomized design in a flat for biotype phenotyping assay. **B** Plants that are susceptible (S) to virulent aphids die during the course of the assay while plants resistant (R) to aphids survive.

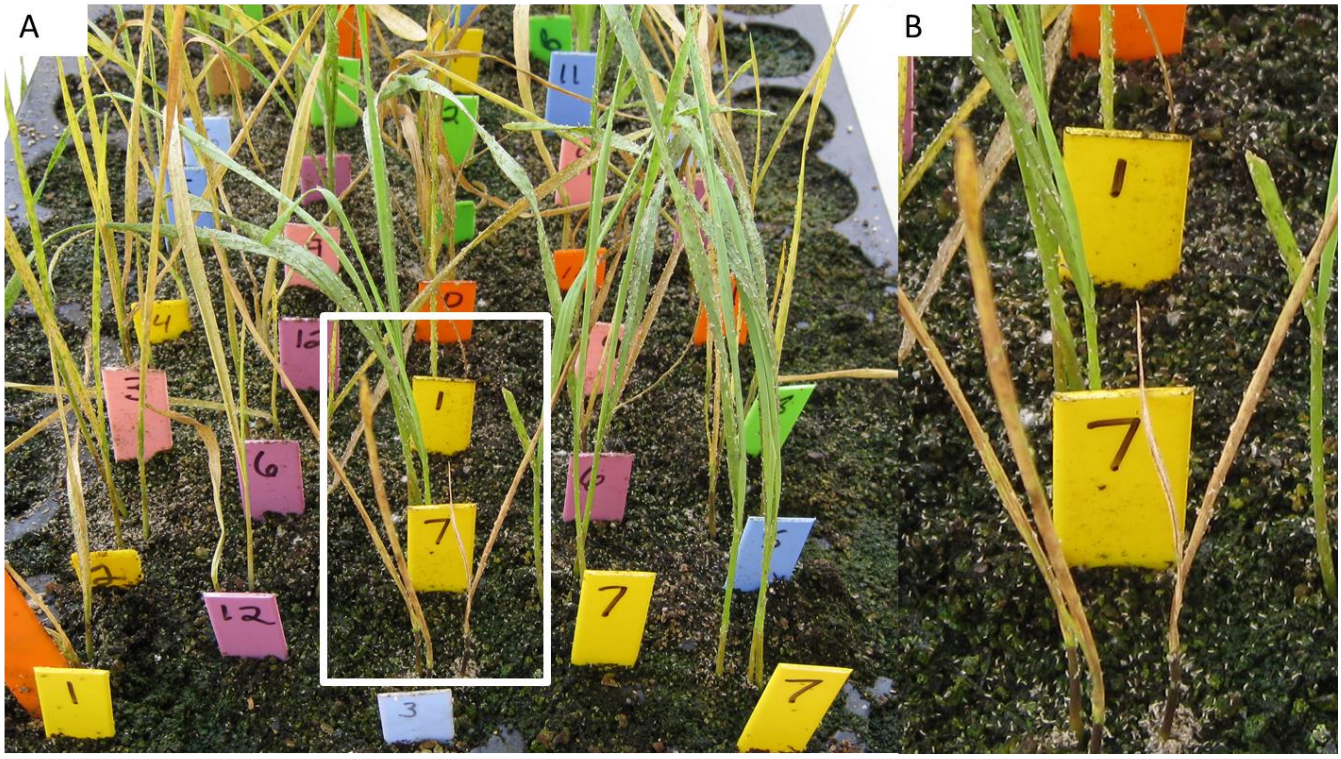


Table 1.1. Novel *Schizaphis graminum* biotypes produced during sexual reproduction break resistance genes in wheat, rye and barley.

	Plant Variety[a]											
				Wheat				Rye		Barley		
Aphid Genotype[b]	Custer	DS 28A (Gb1)	Amigo (Gb2)	CI 17882 (Gb5)	CI 17959 (Gb4)	Largo (Gb3)	GRS 1201 (Gb6)	Elbon	Insave (Gb2, Gb6)	Wintermalt	Post 90 (Rsg1a)	PI 426756 (Rsg2b)
Vector for RPV strain of *Cereal yellow dwarf virus*												
A-3	S	S	S	S	S	S	R	S	R	S	S	S
CC-6	S	R	S	S	S	S	R	R	R	S	R	R
Sg-F (P)	S	R	S	S	S	S	S	S	S	S	R	R
Biotype H	S	S	S	S	S	R	S	S	S	S	S	S
Nonvectors for RPV strain of *Cereal yellow dwarf virus*												
CC-1	S	I	R	S	R	S	S	I	R	S	-	R
K-2	S	S	S	S	S	S	S	S	S	S	R	S
K-3	S	S	S	S	S	S	R	S	R	S	R	S
C-2	S	S	R	S	S	S	R	S	R	S	R	R
BB-1	S	S	S	S	S	S	S	S	R	S	-	R
MM-1	S	S	R	S	S	S	R	S	R	S	R	R
SC (P)	S	S	R	S	S	S	S	S	R	S	R	R

a: Cereal selection, R-gene
b: Aphid genotype, Parental (P) used in the cross to produce the F2 genotypes A3, CC-6, CC-1, K-2, K-3, C-2, BB-1, and MM1.

Reproductive Fitness

Fourth instar nymphs of each *S. graminum* biotype (F, SC and H) were separated in a dish. After 24 h, the recently emerged adults were used to infest the experiment. One adult *S. graminum* was placed in a clip cage attached to one of the plant varieties: barley (cv. Bailey) as a control and the wheat varieties Amigo, Largo, DS28A CI17959, CI17882 and Custer. Three clip cages were attached per plant and infested with the same biotype, performing nine biological replicates of each treatment (wheat line x aphid biotype). Progeny number was counted seven days later. One-way ANOVA was performed grouping the data by biotype or by wheat line. Mean comparisons were performed by Tukey's HSD test.

2-D DIGE

A description of the 2-D DIGE gels was published previously (5) and the dataset was mined for proteins whose expression linked to virus vectoring capacity. Here, we mined the quantitative dataset for proteins that were specifically expressed and/or up-regulated in Biotype H as compared to 11 other *S. graminum* genotypes (Fig. 1.2A). Briefly, Progenesis Samespots v. 3.1 (Nonlinear Dynamics; Newcastle Upon Tyne, United Kingdom) was used to analyze the DIGE images. All images were normalized to an image of a pooled Cy-2-labeled internal standard. Q-value assessment was also performed within the Samespots workflow, and q-values were <0.006 for the entire data set. A correlation analysis in the Samespots statistical module was performed to identify the biotype H-specific protein subset (Fig. 1.2A, Supplemental dataset 1).

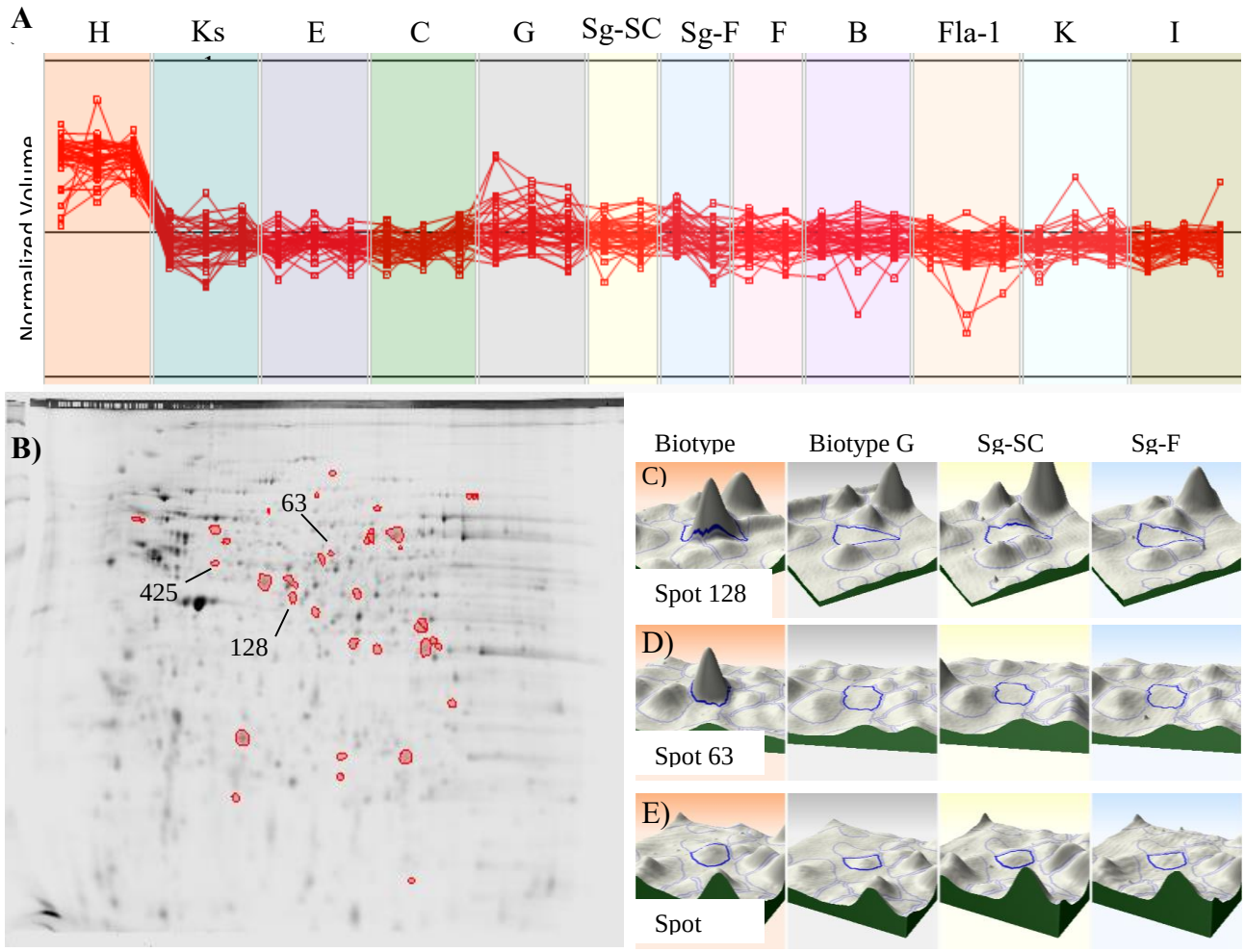
For protein identification, proteins from biotype H were precipitated using 10% trichloroacetic acid in acetone as described (4). Preparative 2-D gels containing 500 µg of non-labeled samples were used for spot picking and mass spectrometry from three biological replicates of *S. graminum* biotype H. The gels were processed as previously described (4). Gels were scanned on the Typhoon Variable Mode Imager (GE Healthcare, Piscataway, NJ) at 100 dpi according to the manufacturer's specifications for CyDyes (GE Healthcare), and Colloidal Coomassie Blue (Invitrogen, Carlsbad, CA) stained gels were visualized with the 632.8 nm helium-neon laser with no emission filter. Spots of interest were manually excised using a 1.5 mm picking pen. In gel digestions were performed as described (6). Extracted peptides were analyzed using LC-MS/MS on Q-ToF Synapt (Waters Corporation, Milford, MA) mass spectrometer as described (4).

LC-MS/MS

Splitless nanoflow chromatography was achieved using a NanoAcquity system (Waters) operating in the vented column configuration. Solvents A and B were 99.9/0.1 water/formic acid and 99.9/0.1 acetonitrile/formic acid, respectively. Trap and analytical columns were prepared in house and packed with 4 µm C12 particles (Phenomenex, Torrance CA, USA). Ten µL of solvent (98% A) were used to flush a 2 µL injection (~1.5 µg of protein) out of a 5 µL loop and onto a self-packed capillary trap column (100 µm ID × 4 cm). After the wash, the six-port valve switched and closed the vent, which initiated the gradient flow (250 nL/min) and data acquisition. Peptides were separated on a 25 cm self-packed column. A 90 min analysis was performed in which solvent B ramped from 2-34 % B over 60 mins (2-62 min); held constant at 80% for 5 mins (63-68 mins), ramped down to initial conditions (68-70) and re-equilibrated for the final 20 mins. (70-90 mins.). Electrospray ionization was initiated by applying a 2.2 kV potential via a liquid junction pre column.

For mass spectrometric analysis, an Orbitrap-Velos (ThermoFisher, Bremen, Germany) was operated in data dependent mode where the top 10 most abundant ions were selected for tandem MS per precursor scan. For MS1 analysis performed in the Orbitrap, a scan range of *m/z* 400-1400 with a resolving power of 60,000 at *m/z* 400 was employed. Automatic gain control was set to 1,000,000 charges with a max ion injection time of 200 ms. For data dependent MS2 scans, performed in the ion trap, an AGC of 8,000 charges and a max ion injection time of 80 ms was used. A 60 s exclusion window was used to avoid repeated interrogation of abundant ions.

Figure 1.2. 2-D DIGE data for protein spots differentially expressed in biotype H. **A** Correlation analysis groups 35 protein spots as differentially expressed in biotype H compared to other *S. graminum* biotypes, ANOVA p-value <0.05. **B** Image of 2-D SDS-PAGE gel containing a Cy-2 labeled, pooled internal standard shows high-resolution separation of protein spots ranging in molecular weight from 200 kDa to less than 20 kDa with isoelectric points between pH 3-10.



Database Searching

Tandem mass spectra were converted into mascot generic format (MGF) peak list files using tools in the Trans-Proteome Pipeline (33). The MGF files were submitted to Mascot v.2.3 (Matrix Science, Boston, MA) (34) for database interrogation. Initial searches against a database containing aphid sequences and common contaminant proteins showed that sample contamination with keratin and other contaminant proteins was not problematic. The data were then searched against a custom database containing the predicted protein sequences from *Acyrthosiphon pisum* and *Buchnera* (download date March 7, 2012). Carbamidomethylation was considered as a fixed modification on cysteine. Oxidation of methionine and deamidation of residues asparagine and glutamine, N-terminal acetylation, and phosphorylation of serine and threonine were also considered as variable modifications. The mass measurement accuracy was set to 30 parts per million (ppm) for precursor ions and 0.8 Daltons (Da) for fragment ions. Peptide assignments from Mascot with an expect value (E-value) less than 0.05 and a precursor mass measurement accuracy of less than 10 ppm were accepted. For spectral counting, data were imported into Scaffold, Version 4 (Proteome Software) and the Fisher's exact was applied to test for differences. Peptides and proteins were filtered using a 99% protein threshold cutoff and a 95% peptide threshold cutoff, which resulted in an overall 0% false discovery rate at both protein and peptide levels. Raw files are available on www.chorusproject.org. Peptide sequence information for the LC–MS/MS data set can be found in Supplemental Table 1.2.

Buchnera EF-Tu Gene Sequencing

DNA was extracted from aphids using a DNease Blood & Tissue kit (Qiagen). *Buchnera aphidicola* EF-Tu was amplified using PCR, a Clonetech Polymerase 2 PCR kit (Takara) and EF-Tu specific primers: forward 5'-ACTTGCGCACCAGTTCTTTT-3' and reverse – 5'-TTTGCCTTCTCAAACTCAGG-3'. The 1721 base pair PCR products were gel purified and sent for Sanger sequencing at the Cornell University Biotechnology Resource Center on an Applied Biosystems Automated 3730xl DNA Analyzer. Forward and reverse sequencing primers were used in each reaction to validate the polymorphisms in at least two independent sequencing reactions per genotype.

RESULTS AND DISCUSSION

Biotype H exhibits distinct protein expression profiles

Both 2-D DIGE and shotgun proteomics enabled us to identify proteins and protein isoforms up-regulated or specifically expressed in biotype H. Correlation and ANOVA analyses on normalized protein spot volumes measured in the DIGE dataset revealed 35 differentially expressed protein spots as compared to 10 other *S. graminum* biotypes (Table 1.2, Fig. 1.2A, Supplemental Table 1.1). The 35 protein spots were first resolved on a non-linear pH 3-10 gradient gel and in the second dimension ranged in molecular weight from approximately 20 kD to 200 kD (Fig. 1.2B). Spots in this dataset were either up-regulated or specifically expressed in biotype H, as no detectible spot volumes were apparent in the same location on the gel images from the other biotypes (typical spot images shown in Fig. 1.2C). Correlation analysis revealed that no other biotype used in the experiment had a group of uniquely expressed protein spots (not shown). LC-MS/MS

analysis of tryptic peptide mixtures from in-gel digests enabled us to assign protein identifications to each spot, resulting in the detection of 27 proteins from the 35 spots. Twenty-nine spots contained a single abundant protein (Table 1.2) whereas six proteins, HSP70 isoform 2, ACO-2, malate dehydrogenase, diacetyl/L-xylulose reductase, cuticular protein 15, and enolase were identified in multiple spots (Table 1.2), indicating that these were protein isoforms differentially expressed in biotype H.

Table 1.2: Proteins and protein isoforms identified using DIGE-LC-MS/MS that are up-regulated in the virulent biotype H of *S. graminum* and their spectral counts in the LC-MS/MS dataset.

							Spectral counts [f]		
Rank[a]	**Fold-change**	**Protein Identification[b]**	**Taxonomy[c]**	**Accession number**	**Score [d]**	**% Coverage [e]**	**Sg-SC**	**Sg-F**	**Biotype H**
Protein Synthesis and Stability									
9	18.1	translation elongation factor-1 gamma	Aphid	gi|193692982	306	20.5	12	23	47
63	6.6	T-complex protein I subunit beta	Aphid	gi|193599182	401	19.1	ND	4	36
128	3.7	Elongation Factor Tu	*Buchnera*	gi|21672772	653	37	ND	26	151
158	3.1	Heat Shock Cognate Protein 70 Isoform 2	Aphid	gi|193603576	545	19.3	29	149	342
425	1.7				346	14			
443	1.7	protein disulfide isomerase ERp57	Aphid	gi|193713655	181	12	4	36	58
Immune and Stress Response									
5	27.1	outer membrane protein F precursor	*Buchnera*	gi|21672622	227	13	29	36	31
109	4.3	GTP cyclohydrolase I	Aphid	gi|193596464	119	13.3	ND	ND	ND
153	3.2	Sushi domain containing protein	Aphid	gi|193671601	93	2.8	ND	ND	ND
164	3.0	ACOnitase (aco-2)	Aphid	gi|193669199	348	15	2	3	28
405	1.8				290	12.5			
195	2.7	Apoptosis inhibitor 5	Aphid	gi|193688054	80	5.2	ND	ND	7
46	8.2	malate dehydrogenase	Aphid	gi|193664537	93	5.7	15	9	55
157	3.1				247	12.3			
217	2.5				84	2.4			
Carbohydrate Metabolism									
46	8.2	glyceraldehyde-3-phosphate dehydrogenase	Aphid	gi|193688110	85	13	103	104	169
73	5.8	putative diacetyl/L-xylulose reductase	Aphid	gi|54287924	139	13	71	46	74
90	4.9				215	22.1			
396	1.8	fructose 1,6-bisphosphate aldolase	Aphid	gi|52630947	192	12.9	93	101	301
Cuticle proteins									
48	8.0	chitin deacetylase 4	Aphid	gi|193652401	113	4.4	ND	ND	ND
191	2.8	Cuticular Protein CPG12	Aphid	gi|193706873	487	17.4	111	182	119
195	2.7	Cuticular protein 15 from low complexity family	Aphid	gi|193620175	230	10.7	86	150	98
217	2.5				219	8.9			
482	1.6				51	1.6			

331	1.9	Cuticle Protein	Aphid	gi|193647873	122	14.5	ND	ND	ND
Macromolecular Transport									
100	4.7	hnRNP A/B-like protein	Aphid	gi|193580163	119	10.2	114	105	75
226	2.5	StAR-related lipid transfer domain protein	Aphid	gi|193709246	157	9	ND	ND	ND
257	2.3	calcineurin A	Aphid	gi|193617722	391	19.4	105	59	47
Other Metabolism									
100	2.7	purine biosynthesis protein 6 (pur6)	Aphid	gi|193636751	318	25	ND	ND	ND
227	2.5	Transaldolase	Aphid	gi|193652569	404	28	3	47	86
237	2.4	Uridine phosphorylase	Aphid	gi|193713581	243	14.2	ND	ND	ND
268	2.3				207	11.8			
443	1.7	Enolase	Aphid	gi|193669445	311	20	103	146	318
396	1.8	ribose-phosphate pyrophospho-kinase 1	Drosophila	gi|21355239	245	14.1	ND	ND	ND
688	1.4	triosephosphate isomerase	Aphid	gi|240849137	320	11	6	21	40

a: Rank is according to fold-change in the 2-D DIGE dataset.
b: Protein common name, as annotated using BLAST[90].
c: Proteome of aphid and *Buchnera* endosymbiont are co-analyzed. Taxonomy represents the top BLAST match of the LC-MS/MS data and indicates whether protein is of insect or endosymbiotic bacterial origin.
d: Mascot protein score, in the case of a small search, such as from gel spots is the sum of the highest ion score for each distinct sequence. ,
e: Percent coverage indicates the percentage of the protein sequence covered by tandem mass spectrometry.
f: Spectral counts for each of the proteins identified by 2-D DIGE in the larger LC-MS/MS dataset. Eight proteins were not identified (109, 153, 48, 331, 226, 100, 237, 396).

Twenty-seven of the proteins characterized by DIGE-LC-MS/MS were also detected in the shotgun proteomics dataset (Table 1.2). Among the twenty-seven proteins that were detected in both datasets, thirteen up-regulated proteins from the DIGE dataset were also up-regulated in the biotype H shotgun dataset using spectral counting (Table 1.2). Although the magnitude of the change was similar for these proteins using DIGE and LC-MS/MS, the precise fold-changes between DIGE and spectral counting varied (Table 1.2). These results are to be expected, as the DIGE quantification is isoform

specific whereas the spectral counts are an average across all the isoforms of a particular protein. In particular, quantification by DIGE is done at the intact protein level (and hence the isoform-specific information) whereas spectral counting is an inference of protein abundance derived from quantification at the peptide level. An additional twenty-three proteins were observed as enriched in biotype H using LC-MS/MS analysis and quantification by spectral counting (Table 1.3). These were not detected as differentially expressed using DIGE-LC/MS/MS. A possible reason for this is the low resolution of 2-D gels resulting in multiple proteins co-resolving in a single spot, a well-described limitation of 2-D gels and DIGE technology (4, 6, 35). Collectively these data show that both approaches enable us to quantify distinct and overlapping sets of proteins and highlight the power of DIGE to quantify specific protein isoforms.

Table 1.3: Identification of proteins up-regulated in *S. graminum* biotype H using LC-MS/MS analysis.

Protein Identification[a]	Accession Number	Taxonomy[b]	Molecular Weight [c]	Spectral Counts	UniProtKB Putative Function[d]
Protein Synthesis and Stability					
Eukaryotic translation initiation factor 3 subunit B	gi|328721361	Aphid	80 kDa	5	Component of the eukaryotic translation initiation factor 3 complex, involved in protein synthesis and stimulates binding of mRNA and methionyl-tRNAi to the 40S ribosome.
Nucleolar KKE/D repeat protein Nop56	gi|215820614	Aphid	52 kDa	23	60S Ribosomal subunit biogenesis
Eukaryotic translation initiation factor 2 subunit 3	gi|193605889	Aphid	42 kDa	9	eIF-2 functions in the early steps of protein synthesis by forming a ternary complex with GTP and initiator tRNA
Immune and Stress Response					
Prohibitin	gi|187119174	Aphid	30 kDa	33	Mitochondrial function, PI3 signaling
3-ketoacyl-CoA thiolase, mitochondrial	gi|193688054	Aphid	42 kDa	9	Mitigates BNIP3-mediated apoptosis and mitochondrial damage.
Cytochrome b-c1 complex subunit	gi|239790824	Aphid	13 kDa	17	Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is part of the mitochondrial respiratory chain that generates an electrochemical potential coupled to ATP synthesis. The complex couples electron transfer from ubiquinol to cytochrome c.
Isocitrate dehydrogenase	gi|193594238	Aphid	39 kDa	10	Oxidative function during TCA cycle, also binds mitochondrial mRNAs

Succinic dehydrogenase	gi|328724302	Aphid	54 kDa	8	Catalyzes the ATP- or GTP-dependent ligation of succinate and CoA to form succinyl-CoA during the TCA cycle.
Fumarase	gi|193643403	Aphid	55 kDa	12	An enzymatic component of the TCA cycle, catalyzes the formation of L-malate from fumarate.
Serine/arginine rich splicing factor-like	gi|239790824	Aphid	18 kDa	15	Component of the spliceosome involved in pre-mRNA splicing
Proliferating cell nuclear antigen (PCNA)	gi|239788493	Aphid	29 kDa	16	Auxiliary protein of DNA polymerase delta, control of eukaryotic DNA replication, a key role in DNA damage response
Buchnera					
Pts system, glucose-specific IIa component	gi|21622956	*Buchnera*	18 kDa	33	Glucose transport
Enolase	gi|26123311	*Buchnera*		26	Glycolysis activator of plasminogen
Glyceraldehyde phosphate dehydrogenase	gi|1094931	*Buchnera*	36 kDa	8	Key enzyme in glycolysis that catalyzes the first step of the pathway by converting D-glyceraldehyde 3-phosphate (G3P) into 3-phospho-D-glyceroyl phosphate.
Sugar and Fatty Acid Metabolism					
L-xylulose reductase	gi|209969796	Aphid	30 kDa	15	Catalyzes the NADPH-dependent reduction of several pentoses, tetroses, trioses, alpha-dicarbonyl compounds and L-xylulose.
Trifunctional enzyme subunit alpha	gi|328717308	Aphid	82 kDa	21	Fatty acid oxidation, lipid metabolism, coA transferase
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	gi|328713184	Aphid	69 kDa	27	Active toward esters of long-chain and very long chain fatty acids such as palmitoyl-CoA, mysritoyl-CoA and stearoyl-CoA
Enoyl-CoA hydratase, mitochondrial-like	gi|28708752	Aphid	31 kDa	9	Mitochondrial fatty acid synthesis
Transcriptional Regulation					
SWI/SNF complex subunit SMARCC2-like	gi|32872125	Aphid	109 kDa	12	Involved in transcriptional activation and repression of select genes by chromatin remodeling
ATP-dependent RNA helicase WM6-like	gi|193652521	Aphid	49 kDa	27	Binds to salivary gland chromosomes to promote open chromatin structure regulating transcription
Macromolecular Transport					
Synaptobrevin	gi|239793435	Aphid	13 kDa	11	Vesicle traffic between golgi and plasma membrane
Sar-1 protein	gi|52630955	Aphid	22 kDa	18	Small GTPase component of the coat protein complex II (COPII) which promotes the formation of transport vesicles from the endoplasmic reticulum (ER).
Importin subunit alpha 7-like	gi|328713140	Aphid	59 kDa	7	Functions in nuclear protein import as an adapter protein

a: Protein common name, as annotated using BLAST[90].

b: Proteome of aphid and *Buchnera* endosymbiont are co-analyzed, indicating whether protein is of insect or bacterial origin.

c: Predicted molecular weight from the annotated gene sequence

d: Uniprot Knowledgebase functional annotation, http://www.uniprot.org/help/uniprotkb

Proteomic variation in the primary bacterial endosymbiont of aphids

S. graminum harbors a primary, maternally-inherited, bacterial endosymbiont in the genus *Buchnera*. *Buchnera* reside within specialized cells located in the abdomen called mycetocytes (36). In this case, the aphid and the bacteria are wholly dependent on each other for survival and reproduction. *Buchnera's* many functions for the aphid are not completely characterized, but they provide essential amino acids that the aphid is unable to synthesize or obtain from the phloem sap diet (37). Variation in *Buchnera* proteomes has been associated with host switching (38) in *Myzus persicae* and also associated with the ability of *S. graminum* to transmit the RPV strain of CYDV (5, 6). Five highly abundant proteins originating from the proteome of *Buchnera aphidicola*, the primary bacterial endosymbiont of aphids, showed distinct variation in biotype H (Tables 1.2 and 1.3). These proteins include Elongation Factor Tu (EF-Tu), Outer membrane protein F precursor (OMP-F), Phosphotransferase system (PTS) glucose specific IIa component, enolase, and glyceraldehyde phosphate dehydrogenase (GAPDH). *Buchnera* cells in biotype H over-express enolase and PTS IIa compared to other aphid biotypes, with 26 and 33 spectral counts respectively, compared to zero spectral counts detected in the other biotypes (Table 1.3). Both of these enzymes increase availability of sugar in bacterial cells, in particular glucose, by different mechanisms. One possibility is that the *Buchnera* of biotype H have an increased metabolic activity to provide another critical, yet unknown function to biotype H and require more sugar to achieve this function. In other bacterial species, enolase, GAPDH and EF-Tu have been characterized as "moonlighting proteins," i.e., multifunctional proteins that contribute to bacterial virulence (39). Their moonlighting notation is derived from evidence that these proteins have

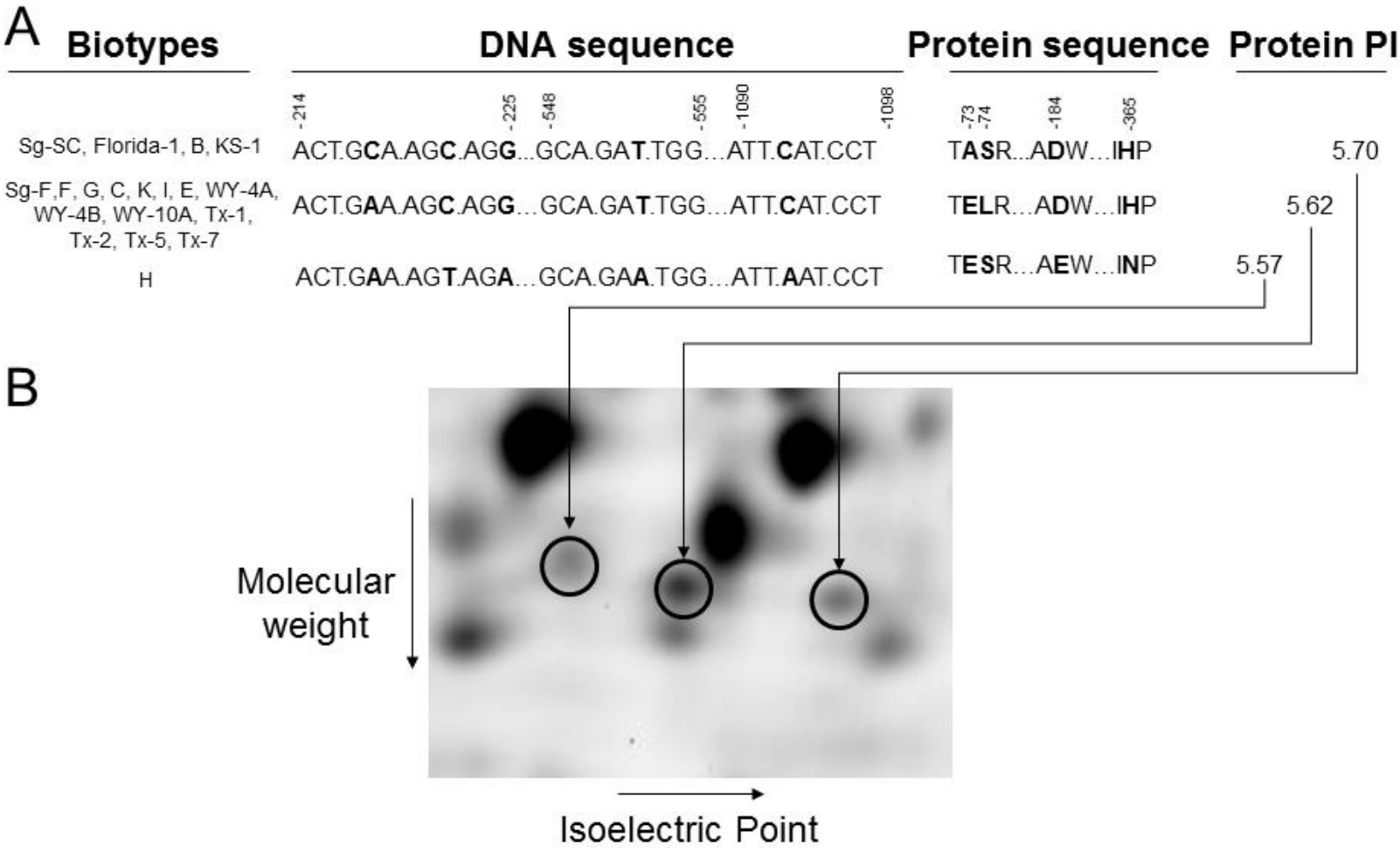
multiple functions outside of their basic metabolic ones and includes both functional and geographical moonlighting within bacterial cells (39, 40). The diversity of isoforms and expression observed in these proteins in the *Buchnera* proteome may correlate with potential moonlighting activities for *Buchnera*.

Facultative bacterial endosymbionts of other aphid species have been implicated in host-based population divergence (41, 42). In *A. pisum*, genotypic variation in the primary endosymbiont *Buchnera,* has been suggested to facilitate such host adaptation. Three isoforms of EF-Tu that varied in charge were detected in *S. graminum*, but each population expressed one major isoform (Fig. 1.3), suggesting that such variation existed within the *S. graminum* populations analyzed here. EF-Tu is a well-known elicitor of the plant pathogen perception system (43, 44), which recognizes microbes via conserved microbial proteins known as pathogen associated molecular patterns (PAMP). It is unlikely that EF-Tu is functioning as a PAMP in the aphid to trigger an immune response but rather may be a signal for a cooperative interaction. Aphids, along with all other hemimetabolous insects with genome sequence information available, have a reduced immune system protein repertoire to combat bacterial infection (3). In particular, they lack the immune deficiency pathway, an adaption that may be necessary for continued symbiosis with *Buchnera*. Nevertheless, we hypothesized that genetic diversity in the primary bacterial endosymbiont of aphids was responsible for the proteomic variation. Such variation may play a role in the expression of the “biotype” phenotype, for example different genotypes of *Buchnera* might confer host-specific advantages or adaptations to the different *S. graminum* biotypes.

To test the source of proteomic variation in *Buchnera*, we sequenced the EF-Tu gene from 19 *S. graminum* biotypes, including several lab-reared strains and those collected from various locations throughout the United States. Three distinct EF-Tu alleles were detected within the populations (Fig. 1.3A). *In silico* translation of the sequenced gene products produced three proteins with distinct isoelectric points (Fig. 1.3). These data corresponded to the detection of three EF-Tu protein isoforms in the *S. graminum* populations that were surveyed using 2-D DIGE (Fig. 1.3B). Gene sequencing revealed the EF-Tu polymorphisms found in biotype H are unique (Fig. 1.3A) and not detected in any other *S. graminum* biotype. These results provide evidence that genetic variation exists in proteins with well-characterized moonlighting activities in other systems, and support the growing body of evidence that variation in endosymbiont populations within aphids imparts adaptive plasticity in host specialization and plant-insect interactions.

Figure 1.3. There is a genetic basis for proteome variation in *Buchnera* endosymbionts of *S. graminum*. Three different isoforms of EF-Tu are expressed by *Buchnera* in *S. graminum.* **A** Differences in the protein isoforms are due to genetic diversity in the *Buchnera* EF-Tu locus. Biotypes Sg-SC, Florida, B, and KS-1 express an isoform with pI 5.7, Biotypes F, G, C, K, I, E, WY-4A, WY-4B, WY-10A, Tx-1, Tx-2, Tx-5 and Tx-7 express a slightly more acidic isoform at pH 5.62. Biotype H is the only one that possesses the isoform at 5.57. *Buchnera* EF-Tu gene sequences detected within the populations of *S. graminum* (A) correspond to the detection of three protein isoforms using 2-D DIGE (B). All three isoforms are resolved on a 2-D gel in the Cy-2 labeled, pooled internal standard containing an equimolar mixture of all the protein extracts used in the experiment. Biotype abbreviations: Wyoming (WY), Texas (Tx), Kansas (Ks), and South Carolina (SC).

Fig. 1.3.


A
Biotypes
DNA sequence
Protein sequence
Protein PI
-214
-225
-548
-555
-1090
-1098
-73
-74
-184
-365
Sg-SC, Florida-1, B, KS-1
ACT.GCA.AGC.AGG...GCA.GAT.TGG...ATT.CAT.CCT
TASR...ADW...IHP
5.70
Sg-F,F, G, C, K, I, E, WY-4A, WY-4B, WY-10A, Tx-1, Tx-2, Tx-5, Tx-7
ACT.GAA.AGC.AGG...GCA.GAT.TGG...ATT.CAT.CCT
TELR...ADW...IHP
5.62
H
ACT.GAA.AGT.AGA...GCA.GAA.TGG...ATT.AAT.CCT
TESR...AEW...INP
5.57
B
Molecular weight
Isoelectric Point

Aphid proteome variation in key regulatory pathways

The aphid proteins that are up-regulated in biotype H from both analyses combined can be categorized broadly into six functional classes. These include (1) protein synthesis and stability, (2) transcriptional regulation, (3) macromolecular transport, (4) cuticle proteins, (5) metabolism (including carbohydrate and fatty acid), and (6) immune and stress response. Interactions between plants and aphids have been compared to those between plants and pathogens (45, 46). In this model, aphids may have a mechanism to detoxify plant defense compounds and cope with the stress placed upon them by the plant immune system. Intriguingly, evidence in the literature supports proteins in all of these classes being involved in coping with stress on the cellular level, in particular oxidative stress. The first signals produced in plants in response to aphid feeding are reactive oxygen species (ROS). Since biotype H is highly successful in colonizing most aphid resistance cereal germplasms (Table 1.1 and (17)), the aphid proteins up-regulated in biotype H may be involved in coping with oxidative stress and ROS detoxification. Although a number of hypotheses are always possible to describe the differential expression of proteins discovered in a comparative proteomics analysis, we will develop our results and discussion, where appropriate, in support of this hypothesis.

Protein synthesis and stability. Among the numerous proteins up-regulated in biotype H involved in protein stability and folding (Tables 1.2 and 1.3), ERp57 and HSP70 have well-described roles in helping cells cope with stress. ERp57 is a soluble thiol-disulfide oxidoreductase localized to the endoplasmic reticulum

(47). A major function for ERp57 is in the folding of glycoproteins; however, ERp57 has been shown to localize to a number of different subcellular organelles to exert multiple functions, including signal transduction and DNA repair (48). The role of ERp57 in insects is almost completely unknown, although in the silkworm *Bombyx mori*, there is evidence that expression of ERp57 may be tied to coping with stress. Expression ERp57 is down-regulated when the silkworms are transferred to a nutrient-poor diet (49). HSP70 is a chaperone protein central to a wide-variety of protein folding processes. In biotype H, two isoforms of HSP70 were detected as differentially expressed in biotype H (Table 1.2), although one isoform was more abundant than the other, 3.1 vs. 1.7 fold-change (Table 1.2). In insects, a role for HSP70 in the protection of insects against oxidative stress has recently been described (50). HSP70 overexpression in *Drosophila melanogaster* hemocytes promoted higher longevity when the insects were exposed to oxidant stress (50). Furthermore, HSP70 overexpression was also linked to an overall decrease in ROS levels in whole flies (50).

Transcriptional regulation. Two proteins involved in chromatin remodeling were observed to be up-regulated in biotype H. SWI/SNF related, matrix associated, actin dependent regulator of chromatin (SMARCC2) is a component of the SWI/SNF complex. These proteins are ATP-dependent chromatin-remodeling factors that regulate eukaryotic transcription. The roles for the SWI/SNF complex in insect development have been well described using *D. melanogaster* (51–58) and are primarily in regulation of homeotic gene expression. In yeast, damage to the extracellular matrix causes transcriptional changes in a

SWI/SNF-dependent manner (59). RNA Helicase WM6, known as UAP56 in *D. melanogaster*, is a ubiquitously expressed DEAD-box protein. It has been shown to function as part of a protein complex involved in transposon silencing (60). These proteins may play a role in differentially regulating genes in biotype H.

Macromolecular transport. Six proteins involved in macromolecular transport were found to be up-regulated in biotype H. Five out of six were discovered in the LC-MS/MS analysis (Tables 1.2 and 1.3). One, the StAR-related lipid transfer protein, was discovered only using DIGE-LC/MS/MS (Table 1.2). The StAR-related protein we found here is a homolog of the *D. melanogaster* ceramide transfer protein (CERT), with 65% similarity and 49% identity between the two proteins. Sphingolipids provide structural integrity to the plasma membrane. Ceramide phosphoethanolamine is a structural analog and functional substitute of sphingomylen in the *D. melanogaster* plasma membrane and is crucial for signaling across the plasma membrane (61). A *D. melanogaster* CERT mutant shows increased plasma membrane fluidity, enhanced susceptibility to oxidative stress and a global increase in the oxidative modification of the *D. melanogaster* proteome (61). Up-regulation of synaptobrevin, sar1, and importin-alpha (Table 1.3) indicate that multiple steps in intracellular transport (from endoplasmic reticulum to golgi, golgi to plasma membrane, nuclear import) are also differentially regulated in biotype H as compared to other biotypes. Biotype H is the most efficient vector for five species of YDV among all tested biotypes and genotypes of *S. graminum* (5). Alterations in intracellular transport pathways may help to explain this observation.

Cuticle proteins. Insects synthesize a cuticle that provides support to the muscles of the insect and protects against environmental stress, predation, and disease. The cuticle is made up of chitin, a biopolymer of N-acetylglucoseamine. Four proteins involved in cuticle biology were detected as differentially expressed in the 2-D DIGE-LC-MS/MS analysis, annotated as chitin deacetylase 4, cuticular protein CPG12, Cuticlular protein 15, and cuticle protein. Three isoforms of cuticular protein 15 were detected on the DIGE gels (Table 1.2), and this protein was also detected in the LC-MS/MS analysis in Sg-F, SC, and biotype H (Table 1.2); however, differences in spectral counts were not significant among biotypes. Thus, certain cuticular protein 15 isoforms in biotype H are slightly up-regulated (2.7, 2.5, and 1.6 fold-change increase for each isoform) but the overall level of this protein is constant in different *S. graminum* biotypes. Intriguingly, the chitin deacetylase 4 protein (or an isoform of this protein) was observed to be up-regulated 8-fold in biotype H in the DIGE analysis (Table 1.2). Chitin deacetylase catalyzes the hydrolysis of acetamido groups of N-acetyl-D-glucosamine in chitin (62). Although the biological roles of chitin deacetylase 4 in aphids are not known, Blast analysis indicates it is homologous to the chitin deacetylases found in other insects and fungi (Supplemental Fig. 1.1). Increased chitin deacetylase in biotype H may reflect a unique role for this protein during embryo production in biotype H: hemimetabolous insects, such as aphids, undergo embryonic molting (63) and chitin deacetylase may be required for this process. Alternatively, there may be a role for chitin deacetylase in protecting the aphid against ingested plant lectins, chitinases and hydrolases that are expressed in the phloem tissue. In the soybean

aphid, *Aphis glycines*, a chitin synthase gene was found to be expressed in many tissues, including the gut (64). In *Trichoplusia ni*, chitin deacetylase is also highly expressed in the insect gut and specifically in response to feeding (65). The acetylase activity of this enzyme alters the structure of chitin in a way that protects it against breakdown by plant enzymes (reviewed in (66)) or buffers the aphid against ingested plant phloem lectins (67). Parallels can be drawn to a similar molecular interplay that has been observed in studies of fungal hyphae invading plant tissues. One way that fungal pathogens evade plant antimicrobial defenses is by enzymatic modification of the invading hyphae to make them less than ideal substrates for plant enzymes. Specifically, exposed fungal chitin polymers are partially de-N-acetylated during infection to escape breakdown by plant hydrolases (68).

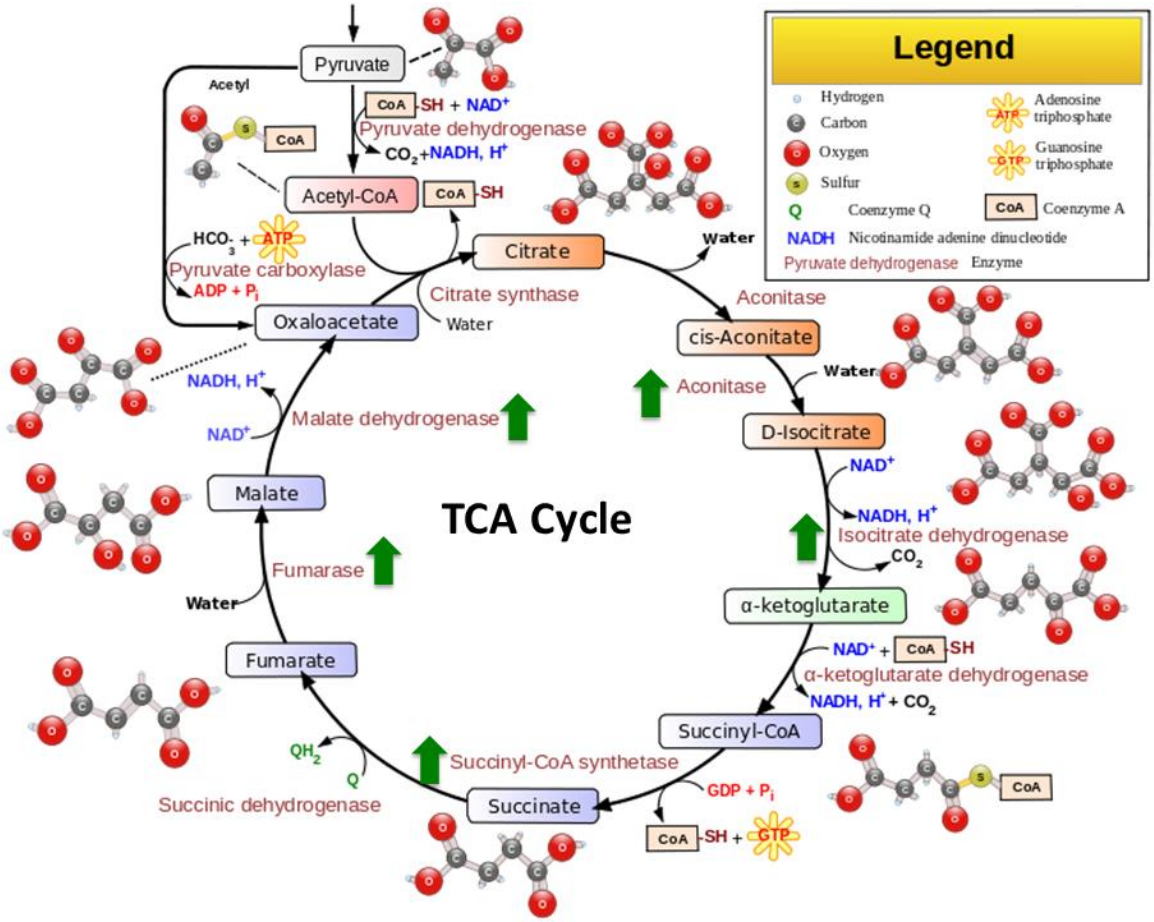
Metabolism. Ten proteins with functions in aphid metabolism were identified as slightly up-regulated in biotype H by DIGE or spectral counting. The pur6 protein was found to be up-regulated by 2.7-fold in biotype H (Table 1.2). This protein functions in the purine biosynthesis as a carboxylase for purine ribonucleotides. In aphids, purine metabolism is regulated by genetic complementarity in the aphid and *Buchnera* genomes (69); however, no changes in *Buchnera* purine metabolic enzymes were observed in biotype H as compared to the other biotypes. Di-acetyl/L-xylulose reductase was identified in the DIGE dataset as two up-regulated isoforms in biotype H (Table 1.2). A second di-acetyl/L-xylulose reductase protein was also identified in the LC-MS/MS dataset as up-regulated in biotype H, with 15 spectral counts (Table 1.3) and these are 70% identical to each other at the amino

acid level (Supplemental Figure 1.2). Di-acetyl/L-xylulose reductase is a multifunctional enzyme with roles in carbohydrate metabolism and detoxification. It reduces high methoxy compounds and converts L-xylulose into xylitol (70). In aphids, this enzyme is up-regulated in response to heat and radiation stress (71) and predation by parasitoids (72). It may also play an important role in plant-*S. graminum* interactions. Indeed, plant polysaccharides differentially influence aphid probing and feeding behaviors. In artificial diet experiments, acetylated pectins are a strong deterrent to *S. graminum*, but not *M. persicae* or *A. pisum*, probing; however, *S. graminum* does eventually feed on the artificial diet containing the acetylated pectins (73). A plausible and testable hypothesis is that expression of di-acetyl/L-xylulose reductase is involved in the reduction of high methoxy plant polysaccharides. Its differential regulation may contribute to adaptation of *S. graminum* to specific host plants.

Immune and stress response. Thirteen aphid proteins involved in immune system and stress response were identified in the combined DIGE and LC-MS/MS analysis. Five of these proteins, aconitase, isocitrate dehydrogenase, succinyl-coA synthetase, fumarase, and malate dehydrogenase are enzymes involved in various steps of the mitochondrial tricarboxylic acid (TCA) cycle (Fig. 1.4), also known as the citric acid or Krebs cycle, were found to be up-regulated in biotype H (Tables 1.2 and 1.3, Fig. 1.4). The TCA cycle produces cellular ATP for energy and NADH during aerobic respiration. There are multiple lines of evidence that pinpoint the TCA cycle as playing a central role in protection against oxidative stress (reviewed in (74)). Oxidative stress specifically modifies aconitase in

houseflies (*Musca domestica)*, causing a loss in catalytic activity and a decrease in lifespan (75). α-ketoglutarate, produced during the TCA cycle, acts as a ROS scavenger and detoxifies H_2O_2 and O_2^- (74), intriguingly also in prokaryotic cells. In fact, modulation of the expression of these mitochondrial enzymes by natural selection has been previously proposed to be involved in evolutionary adaptation and even speciation (76). A hypothesis for the virulence observed in biotype H is that modification of these mitochondrial enzymes, and hence the TCA cycle, helps to enable biotype H to overcome most plant defenses directly mediated by ROS signaling to the aphid upon infestation.

Figure 1.4. The TCA cycle functions in oxidative organisms to drive the production of ATP during oxidative phosphorylation. Schematic of the TCA cycle shows the eight enzymes involved. Five key enzymes in the TCA cycle, malate dehydrogenase, aconitase, isocitrate dehydrogenase, succinic dehydrogenase, and fumarase, are up-regulated and/or display unique isoforms in biotype H. This image has been adapted from http://en.wikipedia.org/wiki/File:Citric_acid_cycle_with_aconitate_2.svg under the Creative Commons Attribution-Share Alike 3.0 Unported License (http://creativecommons.org/licenses/by-sa/3.0/deed.en).



Biotype H presented a reproductive advantage compared to other biotypes

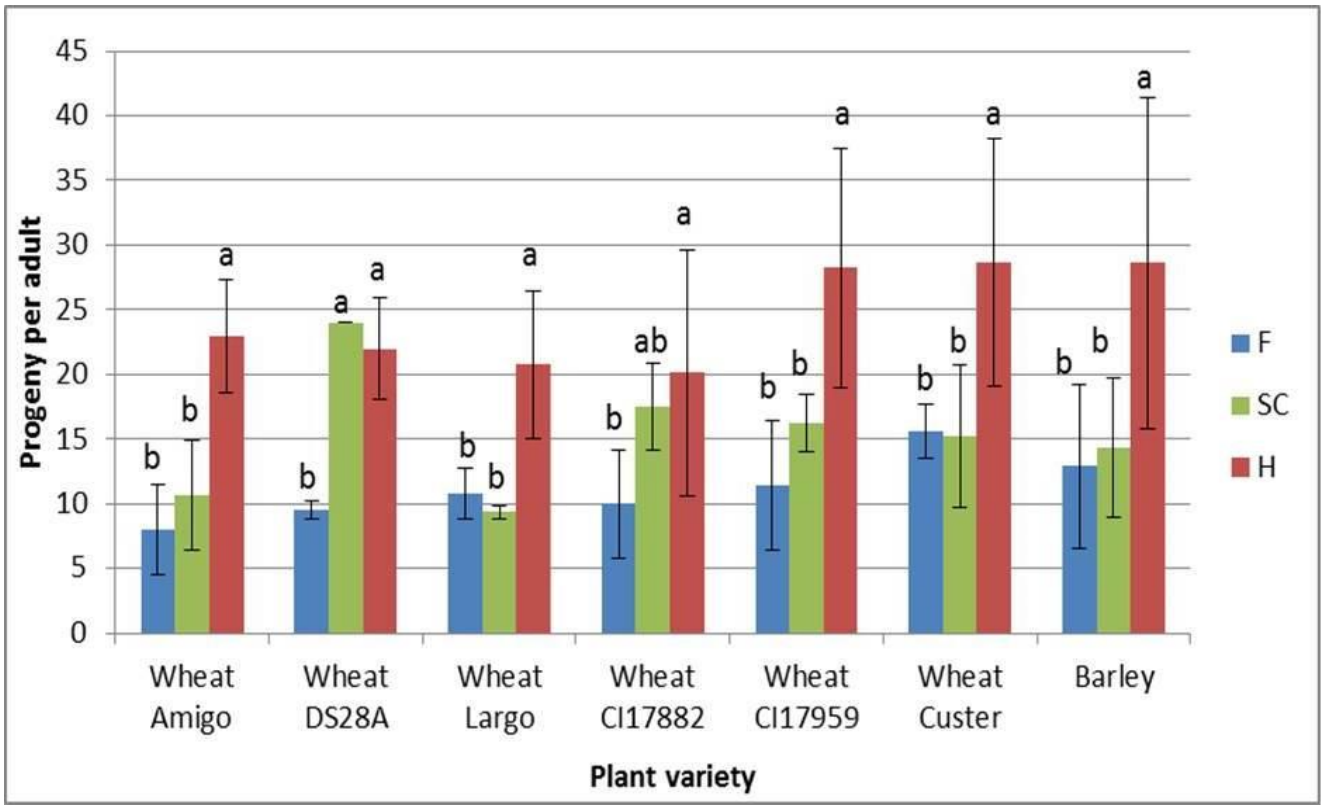
The proteomic data indicate that biotype H is better able to detoxify the plant metabolic weaponry against herbivore attack. If this is a valid hypothesis, one prediction that could be derived is that biotype H would have an increase in reproductive fitness as compared to other biotypes. We tested this hypothesis using biotype H, F, and SC on six wheat germplasms expressing different aphid resistance genes and on one cultivar of barley (Bailey). Progeny means were significantly different between biotypes, $p<0.0001$ (Fig. 1.5). However, the wheat lines did not present a significant effect on progeny means within a biotype, $p=0.107$ (Fig. 1.5). The progeny of biotype H was significantly higher than biotype F on all wheat lines and on barley (Fig. 1.5). When comparing biotype H to SC, the progeny means were not significantly different only for the wheat varieties DS28A and CI17882 (Fig. 1.5). In the virulence assays, we showed that the wheat variety Largo was the only one that survived to biotype H infestation (Table 1.1) but was susceptible to all other biotypes. Interestingly, the reproductive rate of biotype H on wheat Largo was also superior to the other biotypes. These data suggest that Largo was able to recover from the damage caused by biotype H more than from the other biotypes tested in the virulence assays. The proteins identified in the proteomics analysis of biotype H indicate several specific biochemical pathways regulating the aphid side of this interaction. On the plant side, the ability of this wheat line to recover from the infestation of biotype H would be considered as tolerance instead of resistance, because it did not impact aphids' reproductive fitness (no antibiosis or antixenosis effect).

Biotypes F and H are efficient vectors of CYDV-RPV and BYDV-SGV, while biotype SC is a non-vector (5, 31). Our results show that the progeny means of Sg-F and Sg-SC did not differ in any of the plant varieties tested (Fig. 1.5), which suggests that vectoring capacity was not correlated to reproductive fitness when there is no viral infection present. On the other hand, biotype H, which is the most efficient vector of YDVs and the most virulent on agronomic host plants, had a significant reproductive advantage compared to the other two. These data are also supported by the proteomic observations. Previous work examining the proteomic variation underlying vectoring capacity in aphids shows a large non-overlap in the proteins that are linked to vectoring capacity in *S. graminum* (6) and those described here that are found to be associated with host virulence (Tables 1.2 and 1.3). Changes in the biochemical pathways described above are linked with an increased reproductive fitness for biotype H as compared to other *S. graminum* genotypes.

Previous evidence suggested that biotype H might be a cryptic species of *S. graminum* (20). Phylogenetic analysis of the cytochrome oxidase I mitochondrial gene from nine *S. graminum* biotypes showed that biotype H did not fall in any of the three clusters found to classify the other biotypes. Our results show that biotype H has a unique profile of proteins related to virus transmission (5), unique proteins related to host virulence (Tables 1.2 and 1.3) and also unique variants of *Buchnera* proteins (Tables 1.2 and 1.3). The *Buchnera* proteome data suggest that this biotype has a different strain of *Buchnera* and sequencing of EF-TU supports the idea of unique genetic differences in biotype H (Fig. 1.3). During the co-evolution

of aphid populations and endosymbiont strains, *Buchnera* experienced a decrease in its genome size due to the loss of unnecessary genes and retention of genes related to their symbiotic role (77). Additionally, since *Buchnera* is transmitted vertically by aphids and it is located in specific cells in the host, there is little chance of genetic recombination with other bacteria strains (78). Unique polymorphisms, as seen in our results, would more likely be the result of a speciation event as a consequence of aphid-*Buchnera* coevolution. Biotype H cannot produce sexual morphs (79), an observation that is also consistent with our unpublished data. These observations suggest that biotype H has lost its ability to switch between parthenogenesis and sexual reproduction in response to light and temperature cues, which may indicate biotype H is reproductively isolated from other *S. graminum* biotypes in nature. To the best of our knowledge, this study represents the first time insight into the development of a cryptic species has been gleaned from proteomic measurements.

Figure 1.5. *S. graminum* biotypes F, SC and H progeny average numbers and standard deviation of a single female on aphid resistant wheat varieties and barley. Different letters show pairs of means that are significantly different within a plant variety ($p<0.05$).



Novel *S. graminum* biotype profiles are produced via sexual reproduction

The identification of proteomic changes in biotype H of *S. graminum* cannot differentiate between the hypotheses that virulent *S. graminum* are selected by the overuse of aphid resistant plant varieties or whether sexual reproduction of *S.*

graminum generates novel biotypes and environmental or other factors influence their emergence in nature. To test this hypothesis, we subjected an F2 population resulting from a cross between biotype F and SC to biotype virulence phenotyping. In these virulence assays, biotype H and genotype K2 were the most virulent to the agronomic crops (Table 1.1). Except for the wheat variety Largo, all other wheat, rye and barley varieties were susceptible to infestation by biotype H. Largo wheat was the only plant variety to survive biotype H infestation, although it was susceptible to all other *S. graminum* genotypes in the virulence assays. All wheat, rye and barley varieties were susceptible to the genotype K2, except for Post 90 barley (Table 1.1). None of the F2 genotypes that are efficient vectors of CYDV were as virulent as genotype K2, which is a non-vector, confirming that the genes that control ability to transmit virus are by and large not the same as those that are involved in host virulence and ability to overtake plant defenses. Additionally, none of the F2 genotypes shared the virulence profile of the parental genotypes, suggesting that the virulence variability is a multigene trait, and not a gene-for-gene trait, as was thought. In some cases, the plant variety was resistant to both parents, but susceptible to some of the progeny genotypes, as it was the case for the barley varieties Post 90 and PI 426756, carrying *S. graminum* resistance genes Rsg1a and Rsg2b, respectively. On the other hand, the wheat variety GRS 1201, susceptible to both parents, was resistant to many of the F2 genotypes, either vectors or non-vectors. These results prove that there is no correlation in biotype profile and ability to transmit CYDV. Moreover, the data conclusively show that the "biotype" phenotype is generated by genetic recombination in sexual reproduction,

as suggested by some studies (79–81), instead of somatic mutations in clonal populations, as it was hypothesized (82).

S. graminum biotypes have long been classified based on the virulence of their populations to a few *S. graminum* resistant plant varieties, carrying different *S. graminum* resistant genes, thus considering host genetic variation but not aphid genetic variation or co-evolution. For instance, most *S. graminum* biotypes are either susceptible or resistant to both resistant genes in barley, Rsg1a and Rsg2b, carried by different barley varieties, which suggested that the virulence phenotype in aphids in response to these two resistance genes is controlled by one single gene. However, Porter and colleagues (83) found that *S. graminum* biotype TX1 was able to break the resistance of Rsg1a, but not Rsg2b. Our results show that at least two F2 genotypes (K-2 and K-3) are able to break the resistance to one of these genes (Table 1.1) and two are unconfirmed (CC-1 and BB-1) (Table 1.1). These findings prove that aphids vary in their ability to break the resistance conferred by these two genes in barley, suggesting that more than one gene is involved in the aphid response to Rsg1a and Rsg2b.

In line with our current findings, Porter and colleagues (83) proposed that host virulence variability existed naturally within *S. graminum* populations and occurred not exclusively as a consequence of the development of new resistant cultivars, but instead, as a response to various selection pressures, not only on cultivated crops, but mainly on wild grass species. We show here that differentially expressed *Buchnera* and aphid proteins may play a role in the aphids' ability to overcome multiple resistance genes. Many of these proteins are well characterized

to be involved in detoxification, aging, and defense against oxidative damage in other organisms, suggesting for the first time that virulent aphid biotypes are better able to cope with biochemical injury sustained by plant defenses. Adaptation to wild grasses is also correlated with virus vectoring capacity. Populations of aphids adapted to wild grasses are more efficient vectors than populations adapted to cultivated crops (31). Cultivated crops are homogeneous genotypes selected and inbred for pathogen resistance, so the populations of aphids that co-evolved with cultivated crops tend to be less adapted for virus transmission. All these results combined suggest that the emergence of new biotypes is a consequence of co-evolutionary arms race between the aphid, plant host, virus, and endosymbiont, and the differential regulation of proteins involved in response to stress is a molecular signature of this arms race in the aphid.

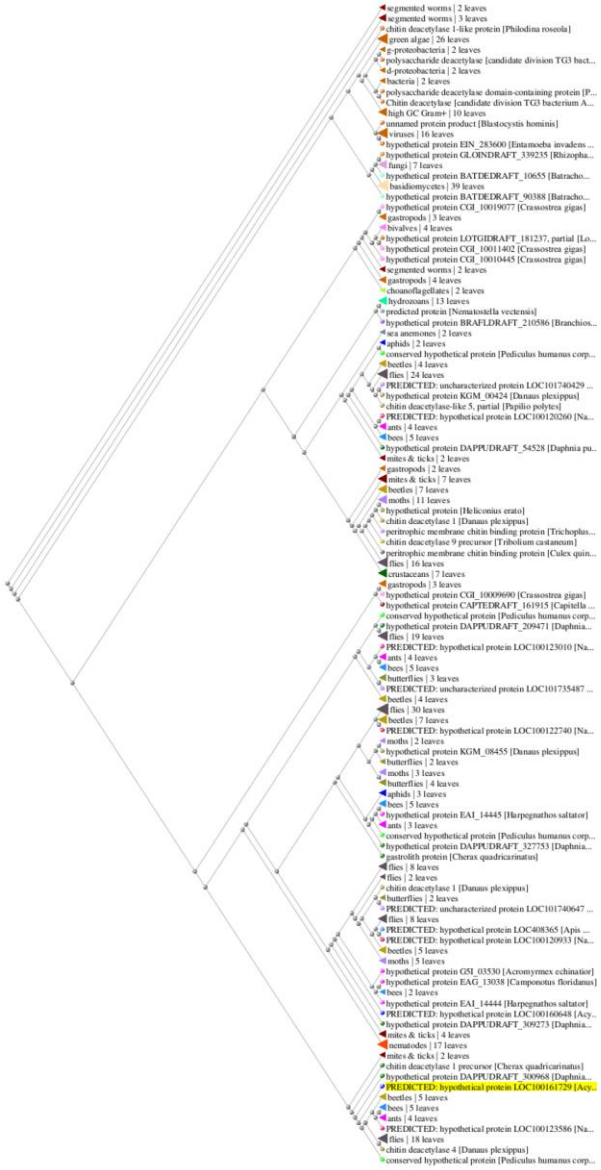
ACKNOWLEDGMENTS

The authors graciously thank Mathieu Rolland (Cornell) for assistance with data analysis, Dawn Smith (Cornell), Jason Ingram (USDA, ARS), and Michelle Sweeney (BTI) for assistance with the aphids, Tara Fish (USDA) for technical assistance, Tom Hammond (Cornell) for care of the plants and Monica Franciscus (Boyce Thompson Institute) for editorial comments on the manuscript. The work was graciously funded by USDA NRI project 1907-22000-018-13, NSF-BREAD project 1109989 and an Embrapa Graduate Student Fellowship to Patricia Pinheiro.

Supplemental Table 1.1. Normalized DIGE intensity values for spots up-specifically regulated in *S. graminum* biotype H

Rank [a]	Anova (p)[b]	Fold[c]	Average Normalised Volumes[d]												
			H	Ks	CC1	E	C	G	SC	Sg-F	F	B	Fla	K	I
4	4.69E-06	33	9.393	0.359	0.643	0.552	0.604	0.716	0.761	0.716	0.717	0.592	0.602	0.647	0.284
5	3.00E-15	27.1	7.249	0.444	0.535	0.649	0.607	0.268	0.454	0.416	0.424	0.387	0.418	0.554	0.585
7	1.09E-07	21.3	7.722	0.435	0.631	0.561	0.479	0.639	0.426	0.574	0.706	0.373	0.363	0.631	0.467
9	6.99E-15	18.1	6.842	0.429	0.386	0.417	0.426	0.694	0.377	0.454	0.448	0.46	0.428	0.413	0.419
46	9.15E-14	8.2	4.652	0.895	0.695	0.82	0.851	0.648	0.923	0.624	0.565	0.661	0.835	0.813	0.831
48	1.08E-05	8	3.907	0.837	1.243	0.507	0.737	0.775	0.93	0.636	0.571	0.732	0.884	0.594	0.485
63	0	6.6	4.715	0.934	0.709	0.958	0.902	0.839	0.918	0.747	0.748	0.754	0.938	0.796	0.908
73	4.78E-10	5.8	3.594	0.737	0.968	0.758	0.924	0.813	0.794	0.857	0.625	0.77	0.922	0.824	0.692
90	2.61E-12	4.9	3.091	0.752	1.127	0.625	0.689	0.824	0.791	0.831	0.737	0.846	0.773	0.669	0.663
100	3.78E-09	4.7	2.95	0.804	0.865	0.781	0.983	0.944	0.848	0.78	0.633	0.746	0.731	1.028	0.799
109	1.04E-06	4.3	3.274	0.839	0.829	0.765	0.761	0.86	1.527	0.931	1.114	1.019	0.825	0.879	0.971
128	2.22E-16	3.7	2.98	0.92	0.94	0.942	0.896	0.923	0.997	0.887	0.896	0.807	0.916	0.887	0.851
153	1.41E-11	3.2	2.253	0.71	0.901	0.877	0.798	0.782	1.17	1.162	1.194	1.052	0.951	0.893	0.843
157	7.75E-09	3.1	2.153	0.704	0.771	0.766	0.798	0.815	0.988	0.993	0.717	0.723	0.708	0.919	0.787
158	8.50E-10	3.1	2.603	0.961	0.952	1.021	0.852	1.019	1.164	1.192	1.137	1.185	1.068	1.049	0.973
164	0.016	3	2.27	0.896	0.964	0.992	0.812	1.235	0.753	0.762	0.818	0.91	0.821	0.892	0.977
178	6.17E-10	2.9	1.99	0.954	0.992	0.749	0.7	0.895	0.686	0.696	0.749	0.901	0.76	0.808	0.715
191	2.91E-14	2.8	2.023	1.008	1.188	0.858	0.81	0.731	1.023	0.824	0.822	1.092	0.979	0.804	0.855
195	6.14E-08	2.7	1.803	0.662	0.715	0.746	0.761	0.943	0.855	0.913	0.982	0.903	0.664	0.806	0.801
217	1.37E-11	2.5	1.924	0.764	0.972	0.862	0.834	0.894	0.93	0.992	0.885	1.043	0.756	0.853	0.805
226	3.09E-06	2.5	1.938	0.823	0.991	0.938	1.039	1.022	0.953	1.017	0.951	0.895	0.781	0.921	0.909
227	7.46E-09	2.5	1.822	0.984	1.038	0.737	0.887	1.057	1.02	1.18	0.963	0.962	0.855	0.863	0.759
236	2.20E-13	2.4	1.878	0.778	0.818	0.965	0.93	0.932	0.903	0.964	0.96	0.944	0.854	1.006	1.012
237	0	2.4	1.964	0.942	0.828	0.927	0.95	1.107	0.911	1.155	1.171	0.958	0.858	0.814	0.867
257	3.75E-11	2.3	1.852	0.86	0.977	0.832	0.807	0.835	1.056	0.928	0.892	0.961	0.926	0.861	0.816
268	1.65E-07	2.3	1.633	1.054	1.204	0.761	0.833	0.874	1.186	0.93	0.825	0.988	0.951	0.73	0.723
276	7.42E-07	2.2	1.495	0.971	0.978	0.968	1.082	1.301	1.087	1.219	1.069	0.998	0.682	0.972	0.965
331	0.013	1.9	1.467	1.186	0.945	0.775	0.82	1.232	1.059	1.181	1.072	0.757	0.884	1.261	0.903
396	2.22E-05	1.8	1.586	1.081	0.966	0.97	1.013	1.039	0.996	0.917	0.881	1.141	0.964	0.927	0.914
405	2.48E-04	1.8	1.509	0.888	0.978	0.965	0.878	1.369	0.852	0.846	0.905	0.893	0.869	0.858	0.947
425	9.70E-05	1.7	1.442	0.839	0.916	0.929	0.989	1.159	0.954	0.92	0.88	0.932	0.828	0.932	0.912
443	1.78E-07	1.7	1.522	1.067	1.137	0.924	0.983	0.913	0.954	1.045	1.069	1.019	0.937	0.908	0.884
482	4.02E-10	1.6	1.367	0.93	1.11	0.953	0.963	1.071	0.967	0.86	0.833	1.061	0.867	0.928	0.958
688	0.031	1.4	1.226	1.032	1.03	0.94	0.929	1.028	0.939	1.048	0.936	0.94	0.879	0.934	0.936
1092	0.153	1.6	1.248	0.869	0.966	0.968	0.895	1.071	0.982	1.088	1.032	0.793	0.858	0.834	0.792

Supplemental Figure 1. Tree-view of BLAST analysis of the aphid chitin deacetylase (gi|193652401) observed to be up-regulated in biotype H shows conservation in diverse taxa such as other insects and even fungi, E-value cut-off 0.05.



Supplemental Figure 2. BLAST alignment of two di-acetyl/L-xylulose reductase proteins identified as up-regulated in biotype H shows 70% identity between the two enzymes.

L-xylulose reductase [Acyrthosiphon pisum]

Sequence ID: ref|NP_001119696.2| Length: 273 Number of Matches: 1

Range 1: 33 to 272 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
345 bits(885)	4e-124	Compositional matrix adjust.	167/240(70%)	197/240(82%)	1/240(0%)

```
Query  1    MEEFFVGKKFIVTGANAGIGETITKRLVQLGAHVFAVGRDPKKLPPAGPNLTPVCADVGD  60
            ME++F GKKFIVTG+ AG+GE IT+RLV LG+ V+AV     +     A PN    +  DV +
Sbjct  33   MEDYFKGKKFIVTGSCAGMGEKITERLVDLGSFVYAVVEKEEGAAKALPNTKQIFCDVSN  92

Query  61   W-NSYDIIKALGPVHGLVNNAGVAFIESFFDMTQEGWDKTLNINARGIVRISQAVAQNMK  119
            W ++Y   +  +GPVHGLVNNAGVA IE FFD+T+ GWDKTLNINAR +VRISQAVA+NM
Sbjct  93   WEDTYKKMYDIGPVHGLVNNAGVAVIEPFFDVTEHGWDKTLNINARALVRISQAVAKNMI  152

Query  120  DAGIKGSIVNVSSTISERAIPDHTSYCASKGAVNQVTRVMSIELGKLGIRTNNVNPTVVM  179
            DAGIKGSIVN+SSTIS RAIPDHT+YCASKGAVNQ+TR M+IELGK GIRTNNVNPTVV+
Sbjct  153  DAGIKGSIVNISSTISTRAIPDHTTYCASKGAVNQITRTMAIELGKYGIRTNNVNPTVVL  212

Query  180  TKMGAKAWSDPEKSNPILSRIPLGRFAECDDVANVTLFLLSDYSTYVNGVSIPVDGGFLA  239
            T+MG  AWSDPEKS PI+ RIPLGRFAE DD+AN  +F+LSDYST VNG ++ VDGGFLA
Sbjct  213  TRMGKIAWSDPEKSGPIMRRIPLGRFAETDDIANAVIFMLSDYSTMVNGTALVVDGGFLA  272
```

REFERENCES

1. **Painter RH**. 1931. Resistance of varieties of winter wheat to Hessian fly Phytophaga destructor. Agricultural Experiment Station, Kansas State College of Agriculture and Applied Science.

2. **Painter RH**. 1951. Insect Resistance in Crop Plants. Macmillan.

3. **International Aphid Genomics Consortium**. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol **8**:e1000313.

4. **Cilia M**, **Fish T**, **Yang X**, **McLaughlin M**, **Thannhauser TW**, **Gray S**. 2009. A comparison of protein extraction methods suitable for gel-based proteomic studies of aphid proteins. J Biomol Tech JBT **20**:201–215.

5. **Cilia M**, **Howe K**, **Fish T**, **Smith D**, **Mahoney J**, **Tamborindeguy C**, **Burd J**, **Thannhauser TW**, **Gray S**. 2011. Biomarker discovery from the top down: Protein biomarkers for efficient virus transmission by insects (Homoptera: Aphididae) discovered by coupling genetics and 2-D DIGE. Proteomics **11**:2440–2458.

6. **Cilia M**, **Tamborindeguy C**, **Fish T**, **Howe K**, **Thannhauser TW**, **Gray S**. 2011. Genetics coupled to quantitative intact proteomics links heritable aphid and endosymbiont protein expression to circulative polerovirus transmission. J Virol **85**:2148–2166.

7. **Cilia M**, **Tamborindeguy C**, **Rolland M**, **Howe K**, **Thannhauser TW**, **Gray S**. 2011. Tangible benefits of the aphid Acyrthosiphon pisum genome

sequencing for aphid proteomics: Enhancements in protein identification and data validation for homology-based proteomics. J Insect Physiol **57**:179–190.

8. **Cilia M**, **Bereman M**, **Fish T**, **MacCoss MJ**, **Gray S**. 2012. Homopteran Vector Biomarkers for Efficient Circulative Plant Virus Transmission are Conserved in Multiple Aphid Species and the Whitefly Bemisia tabaci. J Integr Agric **11**:249–262.

9. **Tamborindeguy C**, **Bereman MS**, **DeBlasio S**, **Igwe D**, **Smith DM**, **White F**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2013. Genomic and proteomic analysis of Schizaphis graminum reveals cyclophilin proteins are involved in the transmission of cereal yellow dwarf virus. PloS One **8**:e71620.

10. **Yang X**, **Thannhauser TW**, **Burrows M**, **Cox-Foster D**, **Gildow FE**, **Gray SM**. 2008. Coupling genetics and proteomics to identify aphid proteins associated with vector-specific transmission of polerovirus (luteoviridae). J Virol **82**:291–299.

11. **Carolan JC**, **Fitzroy CIJ**, **Ashton PD**, **Douglas AE**, **Wilkinson TL**. 2009. The secreted salivary proteome of the pea aphid Acyrthosiphon pisum characterised by mass spectrometry. Proteomics **9**:2457–2467.

12. **Carolan JC**, **Caragea D**, **Reardon KT**, **Mutti NS**, **Dittmer N**, **Pappan K**, **Cui F**, **Castaneto M**, **Poulain J**, **Dossat C**, **Tagu D**, **Reese JC**, **Reeck GR**, **Wilkinson TL**, **Edwards OR**. 2011. Predicted effector molecules in the

salivary secretome of the pea aphid (Acyrthosiphon pisum): a dual transcriptomic/proteomic approach. J Proteome Res **10**:1505–1518.

13. **Wang Y**, **Carolan JC**, **Hao F**, **Nicholson JK**, **Wilkinson TL**, **Douglas AE**. 2010. Integrated metabonomic-proteomic analysis of an insect-bacterial symbiotic system. J Proteome Res **9**:1257–1267.

14. **Cooper WR**, **Dillwith JW**, **Puterka GJ**. 2010. Salivary proteins of Russian wheat aphid (Hemiptera: Aphididae). Environ Entomol **39**:223–231.

15. **Harmel N**, **Létocart E**, **Cherqui A**, **Giordanengo P**, **Mazzucchelli G**, **Guillonneau F**, **De Pauw E**, **Haubruge E**, **Francis F**. 2008. Identification of aphid salivary proteins: a proteomic investigation of Myzus persicae. Insect Mol Biol **17**:165–174.

16. **Nicholson SJ**, **Hartson SD**, **Puterka GJ**. 2012. Proteomic analysis of secreted saliva from Russian wheat aphid (Diuraphis noxia Kurd.) biotypes that differ in virulence to wheat. J Proteomics **75**:2252–2268.

17. **Burd JD**, **Porter DR**. 2006. Biotypic diversity in greenbug (Hemiptera: Aphididae): characterizing new virulence and host associations. J Econ Entomol **99**:959–965.

18. **Porter DR**, **Burd JD**, **Shufran KA**, **Webster JA**. 2000. Efficacy of Pyramiding Greenbug (Homoptera: Aphididae) Resistance Genes in Wheat. J Econ Entomol **93**:1315–1318.

19. **Anstead JA**, **Burd JD**, **Shufran KA**. 2002. Mitochondrial DNA sequence divergence among Schizaphis graminum (Hemiptera: Aphididae) clones from cultivated and non-cultivated hosts: haplotype and host associations. Bull Entomol Res **92**:17–24.

20. **Shufran KA**, **Burd JD**, **Anstead JA**, **Lushai G**. 2000. Mitochondrial DNA sequence divergence among greenbug (Homoptera: aphididae) biotypes: evidence for host-adapted races. Insect Mol Biol **9**:179–184.

21. **Gray S**, **Gildow FE**. 2003. Luteovirus-aphid interactions. Annu Rev Phytopathol **41**:539–566.

22. **Gray S**, **Power A**, **Smith D**, **Seaman A**, **Altman N**. 1991. Aphid Transmission of Barley Yellow Dwarf Virus - Acquisition Access Periods and Virus Concentration Requirements. Phytopathology **81**:539–545.

23. **Aapola AI**, **Rochow WF**. 1971. Relationships among three isolates of barley yellow dwarf virus. Virology **46**:127–141.

24. **D'Arcy C**, **Burnett P**. 1995. Barley yellow dwarf: 40 years of progress. viii + 374 pp.

25. **Miller WA**, **Rasochová L**. 1997. Barley yellow dwarf viruses. Annu Rev Phytopathol **35**:167–190.

26. **Ingwell LL**, **Eigenbrode SD**, **Bosque-Pérez NA**. 2012. Plant viruses alter insect behavior to enhance their spread. Sci Rep **2**:578.

27. **Mauck K**, **Bosque-Pérez NA**, **Eigenbrode SD**, **De Moraes CM**, **Mescher MC**. 2012. Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses. Funct Ecol **26**:1162–1175.

28. **Burrows ME**, **Caillaud MC**, **Smith DM**, **Benson EC**, **Gildow FE**, **Gray SM**. 2006. Genetic Regulation of Polerovirus and Luteovirus Transmission in the Aphid Schizaphis graminum. Phytopathology **96**:828–837.

29. **Burrows ME**, **Caillaud MC**, **Smith DM**, **Gray SM**. 2007. Biometrical genetic analysis of luteovirus transmission in the aphid Schizaphis graminum. Heredity **98**:106–113.

30. **Gray SM**, **Caillaud MC**, **Burrows M**, **Smith DM**. 2007. Transmission of two viruses that cause Barley Yellow Dwarf is controlled by different loci in the aphid, Schizaphis graminum. J Insect Sci Online **7**:1–15.

31. **Gray SM**, **Smith DM**, **Barbierri L**, **Burd J**. 2002. Virus Transmission Phenotype Is Correlated with Host Adaptation Among Genetically Diverse Populations of the Aphid Schizaphis graminum. Phytopathology **92**:970–975.

32. **Gray SM**. 2005. Aphid Transmission of Plant VirusesCurrent Protocols in Microbiology. John Wiley & Sons, Inc.

33. **Deutsch EW**, **Mendoza L**, **Shteynberg D**, **Farrah T**, **Lam H**, **Tasman N**, **Sun Z**, **Nilsson E**, **Pratt B**, **Prazen B**, **Eng JK**, **Martin DB**, **Nesvizhskii AI**, **Aebersold R**. 2010. A guided tour of the Trans-Proteomic Pipeline. Proteomics **10**:1150–1159.

34. **Perkins DN**, **Pappin DJ**, **Creasy DM**, **Cottrell JS**. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20**:3551–3567.

35. **Yang Y**, **Thannhauser TW**, **Li L**, **Zhang S**. 2007. Development of an integrated approach for evaluation of 2-D gel image analysis: impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. Electrophoresis **28**:2080–2094.

36. **Ponsen MB**. 1972. The site of potato leafroll virus multiplication in its vector, Myzus persicae: An anatomical study. H. Veenman.

37. **Baumann P**, **Baumann L**, **Lai CY**, **Rouhbakhsh D**, **Moran NA**, **Clark MA**. 1995. Genetics, physiology, and evolutionary relationships of the genus Buchnera: intracellular symbionts of aphids. Annu Rev Microbiol **49**:55–94.

38. **Francis F**, **Gerkens P**, **Harmel N**, **Mazzucchelli G**, **De Pauw E**, **Haubruge E**. 2006. Proteomics in Myzus persicae: effect of aphid host plant switch. Insect Biochem Mol Biol **36**:219–227.

39. **Henderson B**, **Martin A**. 2013. Bacterial moonlighting proteins and bacterial virulence. Curr Top Microbiol Immunol **358**:155–213.

40. **Henderson B**, **Martin A**. 2011. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infect Immun **79**:3476–3491.

41. **Brady CM**, **Asplen MK**, **Desneux N**, **Heimpel GE**, **Hopper KR**, **Linnen CR**, **Oliver KM**, **Wulff JA**, **White JA**. 2014. Worldwide populations of the aphid Aphis craccivora are infected with diverse facultative bacterial symbionts. Microb Ecol **67**:195–204.

42. **Ferrari J**, **West JA**, **Via S**, **Godfray HCJ**. 2012. Population Genetic Structure and Secondary Symbionts in Host-Associated Populations of the Pea Aphid Complex. Evolution **66**:375–390.

43. **Kunze G**, **Zipfel C**, **Robatzek S**, **Niehaus K**, **Boller T**, **Felix G**. 2004. The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell **16**:3496–3507.

44. **Zipfel C**, **Kunze G**, **Chinchilla D**, **Caniard A**, **Jones JDG**, **Boller T**, **Felix G**. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell **125**:749–760.

45. **Fidantsef AL**, **Stout MJ**, **Thaler JS**, **Duffey SS**, **Bostock RM**. 1999. Signal interactions in pathogen and insect attack: expression of lipoxygenase, proteinase inhibitor II, and pathogenesis-related protein P4 in the tomato,Lycopersicon esculentum. Physiol Mol Plant Pathol **54**:97–114.

46. **Howe GA**, **Jander G**. 2008. Plant immunity to insect herbivores. Annu Rev Plant Biol **59**:41–66.

47. **Frickel E-M**, **Frei P**, **Bouvier M**, **Stafford WF**, **Helenius A**, **Glockshuber R**, **Ellgaard L**. 2004. ERp57 is a multifunctional thiol-disulfide oxidoreductase. J Biol Chem **279**:18277–18287.

48. **Turano C**, **Gaucci E**, **Grillo C**, **Chichiarelli S**. 2011. ERp57/GRP58: a protein with multiple functions. Cell Mol Biol Lett **16**:539–563.

49. **Zhou Z-H**, **Yang H-J**, **Chen M**, **Lou C-F**, **Zhang Y-Z**, **Chen K-P**, **Wang Y**, **Yu M-L**, **Yu F**, **Li J-Y**, **Zhong B-X**. 2008. Comparative proteomic analysis between the domesticated silkworm (Bombyx mori) reared on fresh mulberry leaves and on artificial diet. J Proteome Res **7**:5103–5111.

50. **Azad P**, **Ryu J**, **Haddad GG**. 2011. Distinct role of Hsp70 in Drosophila hemocytes during severe hypoxia. Free Radic Biol Med **51**:530–538.

51. **Curtis BJ**, **Zraly CB**, **Marenda DR**, **Dingwall AK**. 2011. Histone lysine demethylases function as co-repressors of SWI/SNF remodeling activities during Drosophila wing development. Dev Biol **350**:534–547.

52. **Dingwall AK**, **Beek SJ**, **McCallum CM**, **Tamkun JW**, **Kalpana GV**, **Goff SP**, **Scott MP**. 1995. The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. Mol Biol Cell **6**:777–791.

53. **Marenda DR**, **Zraly CB**, **Dingwall AK**. 2004. The Drosophila Brahma (SWI/SNF) chromatin remodeling complex exhibits cell-type specific activation and repression functions. Dev Biol **267**:279–293.

54. **Mohrmann L**, **Langenberg K**, **Krijgsveld J**, **Kal AJ**, **Heck AJR**, **Verrijzer CP**. 2004. Differential targeting of two distinct SWI/SNF-related Drosophila chromatin-remodeling complexes. Mol Cell Biol **24**:3077–3088.

55. **Sun X**, **Morozova T**, **Sonnenfeld M**. 2006. Glial and neuronal functions of the Drosophila homolog of the human SWI/SNF gene ATR-X (DATR-X) and the jing zinc-finger gene specify the lateral positioning of longitudinal glia and axons. Genetics **173**:1397–1415.

56. **Waldholm J**, **Wang Z**, **Brodin D**, **Tyagi A**, **Yu S**, **Theopold U**, **Farrants AKÖ**, **Visa N**. 2011. SWI/SNF regulates the alternative processing of a specific subset of pre-mRNAs in Drosophila melanogaster. BMC Mol Biol **12**:46.

57. **Zeng X**, **Lin X**, **Hou SX**. 2013. The Osa-containing SWI/SNF chromatin-remodeling complex regulates stem cell commitment in the adult Drosophila intestine. Dev Camb Engl **140**:3532–3540.

58. **Zraly CB**, **Marenda DR**, **Dingwall AK**. 2004. SNR1 (INI1/SNF5) mediates important cell growth functions of the Drosophila Brahma (SWI/SNF) chromatin remodeling complex. Genetics **168**:199–214.

59. **Sanz AB**, **García R**, **Rodríguez-Peña JM**, **Díez-Muñiz S**, **Nombela C**, **Peterson CL**, **Arroyo J**. 2012. Chromatin remodeling by the SWI/SNF complex is essential for transcription mediated by the yeast cell wall integrity MAPK pathway. Mol Biol Cell **23**:2805–2817.

60. **Zhang F**, **Wang J**, **Xu J**, **Zhang Z**, **Koppetsch BS**, **Schultz N**, **Vreven T**, **Meignin C**, **Davis I**, **Zamore PD**, **Weng Z**, **Theurkauf WE**. 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell **151**:871–884.

61. **Rao RP**, **Yuan C**, **Allegood JC**, **Rawat SS**, **Edwards MB**, **Wang X**, **Merrill AH**, **Acharya U**, **Acharya JK**. 2007. Ceramide transfer protein function is essential for normal oxidative stress response and lifespan. Proc Natl Acad Sci U S A **104**:11364–11369.

62. **Tsigos I**, **Bouriotis V**. 1995. Purification and characterization of chitin deacetylase from Colletotrichum lindemuthianum. J Biol Chem **270**:26286–26291.

63. **Chapman RF**. 2012. The Insects: Structure and Function5 edition. Cambridge University Press, New York.

64. **Bansal R**, **Mian MAR**, **Mittapalli O**, **Michel AP**. 2012. Characterization of a chitin synthase encoding gene and effect of diflubenzuron in soybean aphid, Aphis glycines. Int J Biol Sci **8**:1323–1334.

65. **Guo W**, **Li G**, **Pang Y**, **Wang P**. 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, Trichoplusia ni. Insect Biochem Mol Biol **35**:1224–1234.

66. **Zhao Y**, **Park R-D**, **Muzzarelli RAA**. 2010. Chitin deacetylases: properties and applications. Mar Drugs **8**:24–46.

67. **Bencharki B**, **Yamani ME**, **Zaoui D**. 2000. Assessment of Transmission Ability of Barley Yellow Dwarf Virus-PAV Isolates by Different Populations of Rhopalosiphum padi and Sitobion avenae. Eur J Plant Pathol **106**:455–464.

68. **El Gueddari NE**, **Rauchhaus U**, **Moerschbacher BM**, **Deising HB**. 2002. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. New Phytol **156**:103–112.

69. **Ramsey JS**, **MacDonald SJ**, **Jander G**, **Nakabachi A**, **Thomas GH**, **Douglas AE**. 2010. Genomic evidence for complementary purine metabolism in the pea aphid, Acyrthosiphon pisum, and its symbiotic bacterium Buchnera aphidicola. Insect Mol Biol **19 Suppl 2**:241–248.

70. **Lee S-K**, **Son LT**, **Choi H-J**, **Ahnn J**. 2013. Dicarbonyl/l-xylulose reductase (DCXR): The multifunctional pentosuria enzyme. Int J Biochem Cell Biol **45**:2563–2567.

71. **Nguyen TTA**, **Michaud D**, **Cloutier C**. 2009. A proteomic analysis of the aphid Macrosiphum euphorbiae under heat and radiation stress. Insect Biochem Mol Biol **39**:20–30.

72. **Nguyen TTA**, **Boudreault S**, **Michaud D**, **Cloutier C**. 2008. Proteomes of the aphid Macrosiphum euphorbiae in its resistance and susceptibility responses to differently compatible parasitoids. Insect Biochem Mol Biol **38**:730–739.

73. **Campbell BC**, **Jones KC**, **Dreyer DL**. 1986. Discriminative behavioral responses by aphids to various plant matrix Polysaccharides. Entomol Exp Appl **41**:17–24.

74. **Mailloux RJ**, **Bériault R**, **Lemire J**, **Singh R**, **Chénier DR**, **Hamel RD**, **Appanna VD**. 2007. The tricarboxylic acid cycle, an ancient metabolic network with a novel twist. PloS One **2**:e690.

75. **Yan LJ**, **Levine RL**, **Sohal RS**. 1997. Oxidative damage during aging targets mitochondrial aconitase. Proc Natl Acad Sci U S A **94**:11168–11172.

76. **Das J**. 2006. The role of mitochondrial respiration in physiological and evolutionary adaptation. BioEssays News Rev Mol Cell Dev Biol **28**:890–901.

77. **Gil R**, **Sabater-Muñoz B**, **Latorre A**, **Silva FJ**, **Moya A**. 2002. Extreme genome reduction in Buchnera spp.: toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci U S A **99**:4454–4458.

78. **Pérez-Brocal V**, **Gil R**, **Moya A**, **Latorre A**, **Pérez-Brocal V**, **Gil R**, **Moya A**, **Latorre A**. 2011. New Insights on the Evolutionary History of Aphids and Their Primary Endosymbiont Buchnera aphidicola, New Insights on the Evolutionary History of Aphids and Their Primary Endosymbiont Buchnera aphidicola. Int J Evol Biol Int J Evol Biol **2011, 2011**:e250154.

79. **Ullah F**, **Peters DC**. 1996. Sexual Reproduction Capabilities of Greenbugs (Homoptera: Aphididae). J Kans Entomol Soc **69**:153–159.

80. **Puterka GJ**, **Peters DC**. 1989. Inheritance of greenbug, Schizaphis graminum (Rondani), virulence to Gb2 and Gb3 resistance genes in wheat. Genome **32**:109–114.

81. **Shufran KA**, **Peters DC**, **Webster JA**. 1997. Generation of clonal diversity by sexual reproduction in the greenbug, Schizaphis graminum. Insect Mol Biol **6**:203–209.

82. **Starks KJ**, **Schuster DJ**. 1976. Greenbug: Effects of Continuous Culturing on Resistant Sorghum. Environ Entomol **5**:720–723.

83. **Porter DR**, **Burd JD**, **Mornhinweg DW**. 2006. Differentiating greenbug resistance genes in barley. Euphytica **153**:11–14.

Chapter 2

Host plants indirectly influence plant virus transmission by altering gut cysteine protease activity of aphid vectors

Research article in Molecular and Cellular Proteomics (2016) papers in press, on December 8, 2016 as Manuscript M116.063495, DOI: 10.1074/mcp.M116.063495

Patricia V Pinheiro [1,2,3], Murad Ghanim [4], Mariko Alexander[5], Ana Rita Rebelo[2], Rogerio S Santos[2], Benjamin C Orsburn[6], Stewart Gray[5,7] and Michelle Cilia[1,5,7,*]

1 Department of Entomology, Cornell University, Ithaca, NY (pvp23@cornell.edu)
2 Boyce Thompson Institute for Plant Research, Ithaca, NY
3 Embrapa Rice and Beans, Santo Antônio de Goiás, Brazil
4 Department of Entomology, Volcani Center, Bet Dagan, Israel
5 Plant Pathology and Plant Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY
6 Thermo Fisher, San Jose, CA
7 USDA Agricultural Research Service, Emerging Pests and Pathogens Research Unit, Robert W. Holley Center for Agriculture and Health, Ithaca, NY

*Corresponding Author:

Michelle Cilia mlc68@cornell.edu; michelle.cilia@ars.usda.gov

ABSTRACT

The green peach aphid, Myzus persicae, is a vector of the Potato leafroll virus (PLRV, Luteoviridae), transmitted exclusively by aphids in a circulative manner. PLRV transmission efficiency was significantly reduced when a clonal lineage of M. persicae was reared on turnip as compared to the weed physalis, a transient effect caused by a host-switch response. A trend of higher PLRV titer in physalis-reared aphids as compared to turnip-reared aphids was observed at 24h and 72h after virus acquisition. The major difference in the proteomes of these aphids was the upregulation of predicted lysosomal enzymes, in particular the cysteine protease cathepsin B (cathB), in aphids reared on turnip. The aphid midgut is the site of PLRV acquisition, and cathB and PLRV localization were starkly different in midguts of the aphids reared on the two host plants. In viruliferous aphids that were reared on turnip, there is near complete co-localization of cathB and PLRV at the cell membranes, which was not observed in physalis-reared aphids. Chemical inhibition of cathB restored the ability of aphids reared on turnip to transmit PLRV in a dose-dependent manner, showing that the increased activity of cathB and other cysteine proteases at the cell membrane indirectly decreases virus transmission by aphids. Understanding how the host plant influences virus transmission by aphids is critical for growers to manage the spread of virus among field crops.

INTRODUCTION

Aphids are small insects that feed exclusively on the phloem sap of plants, causing significant damage to agronomic crops. However, their major economic importance is that they are the most numerous vectors of plant viruses, such as the Poleroviruses in the *Luteoviridae*, which we will refer to collectively as luteovirids in this manuscript. Luteovirids are single stranded, positive sense, non-enveloped RNA viruses that infect a range of economically important crops and weedy hosts. Luteovirids, including *Potato leafroll virus* (PLRV), cause severe yield losses in agronomic crops around the world and are transmitted exclusively by aphids in a circulative manner. Circulative transmission requires a series of spatially and temporally regulated, largely unknown protein interactions with the virus structural capsid proteins (1, 2). There is no cure for viral infection in plants, therefore, the only options are to prevent or avoid infection (3). Host resistance is the ideal method to prevent infection, but despite intensive efforts to identify or breed for resistance, few commercialized luteovirid-resistant cultivars have been released. Controlling aphid vectors using pesticides is costly, and to be effective, information about vector phenology is necessary. Disrupting an aphid's ability to transmit a virus into or within a crop represents a different approach and a promising means by which to control virus spread (3, 4).

Aphids acquire and transmit luteovirids as intact virions, not viral RNA, and there is no evidence to show that luteovirids replicate in their aphid vectors (3, 5). Luteovirids are nonspecifically ingested from the phloem sap together with sap proteins (6) while the aphid is feeding on an infected plant. To be transmitted to a

new plant, luteovirids must overcome physical barriers within the insect, the gut and the accessory salivary glands, a process that is mediated by virus-vector species-specific protein interactions. The virus must first be internalized by gut cells (6–12). Detailed microscopic investigations revealed that the virus moves via endosomes in the gut, with different virus species displaying different affinities to various regions of the gut (i.e., midgut or hindgut). PLRV is acquired into midgut epithelial cells (13). Virions bind to the luminal (apical) plasma membrane, stimulating the formation of coated pits and enter the gut epithelial cells via a receptor-mediated endocytosis mechanism (14). Aphid membrane alanyl aminopeptidase N (APN) has been identified as a cell surface receptor for *Pea enation mosaic virus* (PEMV, genus *Enamovirus*) in the pea aphid, *Acyrthosiphon pisum* (2, 15). Once inside the aphid cell, the virus particles remain in membrane-bound vesicles, during transport through the cytoplasm, and this is universally true for every species of luteovirid studied by microscopy, to date. Unlike when the virus is in plant cells (16), in aphid cells, virions are never observed free in the cytoplasm. The observation that virus-containing tubular vesicles connect to aphid cellular organelles supports hypothesis that the virus is transported intracellularly through the gut endomembrane system. Membrane-bound vesicles containing virions in gut cells of *Myzus persicae* and other aphid species have been observed to connect to lysosomes and lysosomal-like organelles (10, 13). Following transport through the endosome, PLRV and other luteovirids can be observed between the plasmalemma and the basal lamina of the gut epithelia where they are then released into the open circulatory system of the aphid and quickly diffuse (10, 13).

Once the virus reaches the accessory salivary glands, the virus is endocytosed (17, 18), transported through the cells within vesicles, and released into the salivary duct where it can be inoculated into plants together with the saliva as the insects feed. Aphids that acquire luteovirids from an infected plant remain viruliferous for their entire life (5).

Luteovirids promote their own plant-to-plant spread by influencing plant host selection and feeding behavior of the insect vector (19, 20) as well as affecting the production of winged, migratory individuals (21, 22). Aphids are more attracted to plants infected with circulative viruses that they transmit than to healthy plants or to plants infected with viruses that have other modes of transmission. Aphids tend to remain on plants infected with circulative viruses longer than plants infected with cuticle-associated viruses. These findings suggest that different transmission modes shape the extent to which viruses influence their vectors (22–28). Positive or neutral effects on vector performance have been extensively reported for persistently transmitted viruses that are dependent on their insect vectors for transmission (20, 29–36). On the other hand, negative and sometimes neutral effects on insects have been reported mainly for plants infected with viruses and other pathogens that are not transmitted or not exclusively transmitted by the insect species studied (24, 37–41). Collectively, these studies show that viruses have been favored by natural selection to alter vector behavior via controlling vector interactions with their host plants.

Here, we observed that *M. persicae* changed its vectoring ability in a host-dependent manner, and we investigated how the aphid's host plant impacts virus

transmission at the molecular level. A clonal lineage of *M. persicae* reared on a PLRV host plant (physalis) and PLRV non-host plant (turnip) showed significant variation in PLRV transmission efficiency. Using organismal, biochemical, molecular, proteomic, and imaging approaches, we show that high levels of cysteine proteases at the cell membrane in aphids reared on turnip are indirectly responsible for the host-dependent change in the virus transmission phenotype in *M. persicae*.

EXPERIMENTAL PROCEDURES

Insects. Parthenogenic reproducing colonies of the same clonal lineage (genetically identical individuals) of *M. persicae* Sultz (the green peach aphid) were maintained on caged physalis (*Physalis floridana)* or turnip (*Brassica rapa*) at 20°C with an 18-hour photoperiod for a minimum of four months prior to the experiments and proteomics analyses.

PLRV transmission assays. To test the host switch effect on PLRV transmission, aphids from both colonies, turnip (T-Myzus) and physalis (P-Myzus), were transferred to PLRV-infected Hairy nightshade (*Solanum sarrachoides,* HNS) detached leaves for an acquisition access period (AAP) of 24 hours, which is enough time for virus acquisition by aphids (Fig. 1, (42, 43). After collecting the data for this initial experiment, another experiment was performed with an AAP of 48 hours to test whether a longer AAP would increase the virus transmission rate by T-Myzus. HNS seedlings were *Agrobacterium*-infiltrated with the PLRV infectious clone as described (44) and used as the source of virus. After that, 10

aphids were transferred to each healthy potato cv. Red Maria seedling (n=10 plants) for an inoculation access period (IAP) of 48 hours, as described (42, 43). Potato seedlings were treated with imidacloprid to eliminate aphids after the IAP. After three weeks, systemic infection of PLRV was detected in the recipient plants by double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) using α-PLRV antibodies (Agdia).

To test whether the impact of turnip feeding on PLRV transmission was transient, T-Myzus were transferred to physalis plants for three days and then transferred to PLRV-infected HNS for a 48 hr AAP. Similarly, P-Myzus were transferred to turnip plants for three days and then fed to PLRV-infected HNS detached leaves for the same 48 hr AAP. Five aphids were transferred to healthy potato seedlings for the transmission assay for a 48 hr IAP, in 10 replicates per treatment. Three weeks later, the systemic PLRV infection in potato plants was detected by DAS-ELISA using α-PLRV antibodies. The proportion of plants infected with PLRV was compared to non-infected plants using the Chi square test.

Aphid reproduction on different host plants. We measured the effect of each host plant on aphid reproduction and overall weight. Fourth instar nymphs of P-Myzus were transferred to either a turnip or a physalis plant and after 24 hours, nymphs that had molted to adults were transferred to a fresh turnip or physalis plant in four biological replicates to measure the host effect on aphid reproduction. Fifteen days later, the progeny was counted. The individual weight of adults was obtained by averaging the weight of 15-30 adults per replicate. Progeny counts

and adult weights were analyzed by one-way ANOVA, and the means were compared using the Student's t-test.

PLRV acquisition by aphids. P- and T-Myzus were fed on PLRV-infected HNS plants for a 24h AAP. Aphids were then transferred to an artificial diet (45) sandwiched between thinly-stretched parafilm membranes for gut clearing. Aphid cohorts were collected at 24h or 72h after the start of the AAP for PLRV quantification by digital droplet PCR (ddPCR), using the QX100 droplet digital PCR system (Bio-Rad). The ddPCR reaction for PLRV consisted of 10 uL of 2X ddPCR Evagreen SuperMix (Bio-Rad), 1 uL of combined primers at 10 uM each (PLRV CP qPCR F1 5'-CTAACAGAGTTCAGCCAGTGG-3'; PLRV CP qPCR R1 5'-TGTCCTTTGTAAACACGAATGTC-3'), 7 uL of dH2O and 2 uL of DNA diluted at 1:800 in a final volume of a 20 uL reaction. The entire 20 uL reaction was then loaded into a disposable DG8 droplet generator cartridge secured in the cartridge holder (Bio-Rad). A total of 70 uL of droplet generator oil for Evagreen (Bio-Rad) was also loaded into the disposable DG8 droplet generator cartridge. The cartridge holder was placed into the QX100 droplet generator (Bio-Rad) where droplets were generated. Droplets were then transferred to a 96-well plate (Eppendorf) and the plate was sealed with an easy pierce foil seal (Bio-Rad). PCR amplification was carried out on the Applied Biosystems 2720 Thermocycler. The thermocycling conditions started at 95°C for 5 min, followed by 40 cycles of 95°C for 30 sec and 60°C for 1 min, 1 cycle at 4°C for 5 min, 1 cycle at 90°C for 5 min and ending at 12°C. Following amplification, the plate was inserted into the droplet reader cassette and loaded into the droplet reader (Bio-Rad). The droplets were

automatically read at a rate of 8 wells per 15 min. The ddPCR droplet data were analyzed using the QuantaSoft analysis software Version 1.7.4 2014 (Bio-Rad), which presents the target results as copies per µL of PCR mixture. The default settings to analyze the data were as follows: experiment set to ABS; SuperMix: QX200 ddPCR EvaGreen Super; Channel 1: FAM; and Channel 2: Vic/Hex. The digital droplet reader determined the number of copies per µl using the Poisson Distribution calculator. The number of copies obtained in two independent experiments were averaged and analyzed by One-Way Anova.

2-D DIGE. We used gel-based separation and quantification of intact proteins, 2D-DIGE, to measure the relative quantification of protein expression between P-Myzus and T-Myzus. Three biological replicates of T- and P-Myzus (all life stages) were weighed and frozen at -80°C in 50 mL BD-Falcon (Franklin Lakes, NJ, USA) for protein extraction. Care was taken to remove all plant and soil debris from the aphids before freezing, so as not to contaminate the aphid protein samples. Proteins were extracted using a TCA-acetone protein precipitation protocol optimized for 2-D gel electrophoresis of aphid proteins (46). Protein samples were labeled with Cy3 or Cy5 according to the manufacturer's instructions (GE Healthcare; Piscataway, NJ). A Cy2 internal standard containing an equal amount of proteins from all the biological replicates was used for relative quantification by DIGE technology. Cy-dye labeled samples were grouped randomly during 2-D gel electrophoresis so that each gel contained a Cy3 and a Cy5 labeled sample, together with the Cy2-labeled, pooled internal standard. A dye swap was

performed to control for any labeling bias. Analytical gels containing Cy dye-labeled samples were used for quantitative analysis and preparative gels containing non-labeled samples were used for spot picking. A total of 50 µg of each Cy dye-labeled sample or 500 µg of non-labeled protein were loaded onto immobilized pH gradient (IPG) strips (pH 4 to 7, 18 cm; GE Healthcare) during an overnight passive rehydration of the strips according to the manufacturer's specifications for the analytical and preparative gels, respectively. The first dimension was run on the IPGphor II (GE Healthcare) at 20°C with the following settings: step 1: step and hold for 500 V, 1 h; step 2: gradient 1,000 V, 2 h; step 3: gradient 8,000 V, 3 h, and step 4: step and hold 8000V until 34,000 V for a total focusing time of 10 h. Next, the IPG strips were reduced for 15 min with 64.8 mM of dithiothreitol in SDS equilibration buffer (50 mM Tris-HCl [pH 8.8], 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue) and then alkylated for 15 min with 135.2 mM of iodoacetamide in SDS equilibration buffer. The second dimension was carried out using 8-15% gradient tris-glycine gels (Jule, Inc, Milford, CT). Gels were cast 1mm thick by 25.5 cm wide by 20.2 cm tall with an acrylamide:bis ratio of 38:1. The Ettan DALT Six system (GE Healthcare) was used to run the second dimension at 25°C with the following settings: step 1, 10 mA/gel, 1 h; step 2, 40 mA/gel, 6 h or until the bromophenol blue front ran to the bottom of the gels. The preparative gels were fixed in a solution of 15% methanol, and 7.5% acetic acid for one h, stained overnight in Colloidal Coomassie Blue (Invitrogen) and destained in water for 12 h prior to spot picking.

Gel analysis. Gels were scanned on the Typhoon Variable Mode Imager (GE Healthcare) at 100 dpi according to the manufacturer's specifications for Cy dyes (GE Healthcare) and Colloidal Coomassie Blue (Invitrogen) stained gels were visualized with the 632.8 nm helium-neon laser with no emission filter. DIGE gel images were analyzed using Progenesis Samespots v. 3.1 (Nonlinear Dynamics; Newcastle Upon Tyne, United Kingdom). Fifty manual alignment seeds were added per gel (~12 per quadrant) and the gels were then auto-aligned and grouped according to host plant for analysis. Spots were selected as being differentially expressed if they showed greater than a 2.0 fold change in spot density and an ANOVA P-value of <0.05.

Mass spectrometry. Gel plugs were picked from the preparative gels and proteins were prepared for mass spectrometry as described previously (46). Peptides were analyzed using a Q Exactive (Thermo Fisher Scientific) mass spectrometer. For analysis on the Q Exactive, 5 µl of the in gel digest was loaded onto a 5 cm C-18 PepMap trapping column with an EasyNanoLC 1000 (Thermo Fisher Scientific) system. The peptides were loaded at a maximum pressure of 280 bar and washed with 3 µL of 0.1% formic acid (FA) at the same pressure settings. The peptides were separated by a 20cm hand packed and pulled C-18 PepMap column where the voltage for ionization was applied at a liquid junction prior to the column. A 30 min gradient was applied, beginning at 4% acetonitrile (ACN), 0.1% FA and ramping linearly to 35% ACN, 0.1% FA. The remaining time was used for a high organic and re-equilibration of the trapping and analytical columns. Nanospray ionization was achieved with a NanoFlex ion source (Thermo Fisher Scientific)

operating at 1.9 kV and an ion transfer tube temperature of 275°C. MS/MS analysis was performed using software version 2.2 SP1. Profile MS1 spectra were obtained from an *m/z* of 400 to 1600 at 140 k resolution with an automatic gain control (AGC) target of $3.0e^6$ charges and a maximum fill time of 50 ms. The 20 most intense ions from each MS1 scan with charge states of 2-7 were selected for fragmentation with preference given to those with an isotopic distribution matching that of average using the peptide match option “preferred”. Selected ions were isolated in a 1.6 Da window centered on the most intense isotope with an AGC target of $1e^5$ and a maximum fill time of 50 ms. Fragmentation was normalized to a collision energy of 27 for each selected ion. Targeted ions were then placed on the dynamic exclusion list for 15 seconds to allow multiple fragmentations of each ion. All raw files were converted to Mascot generic format (mgf) using Proteome Discoverer v. 1.3. For searching the 2-D DIGE data, a custom database was built containing *M. persicae* proteins sequences available from Aphidbase and available *Buchnera* sequences containing a total of 2,736 proteins. MGF files were searched using Mascot v. 2.5 with oxidation of methionine and deamidation of glutamine and asparagine as variable modifications using trypsin as the enzyme and three missed cleavages allowed. No fixed modifications were used for the gel spot searches. Mass error tolerances were set at 20ppm for the precursor ion and 0.02 Da for the product ions. An expect score of 0.05 or less was used as a threshold cut-off for peptide identification. As is typical for gel spot analysis, no false discovery rate is provided as the data set for each gel spot is too small for FDR estimation (46).

Differential expression of cathepsin B in aphids was verified using relative peptide quantification by Selected Reaction Monitoring mass spectrometry (SRM). Nano-flow liquid chromatography was performed using an Easy nLC 100 (Thermo Fisher Scientific) in a vented-tee configuration. A 10 cm trapping column (100 µm I.D. x 360 O.D.) and a 15 cm analytical column (75 µm I.D. x 360 O.D.) were packed with 3 µm C18 reverse phase particles (CorConnex). Emitter tips (New Objective) were trimmed to 4 cm. Two µL of the 1 µg/mL digested aphid protein extracts were loaded onto the trapping column and eluted with a flow-rate of 300 nL/mn. The gradient ramped from 5% B (95:5 acetonitrile/formic acid) to 37% B across 110 min, and then increased to 80% B and held constant for 5 min. Electrospray ionization (ESI) was initiated using a CorConnex plug and play nanoLC-ESI interface applying 1.4 kV via a liquid junction distally from the ESI tip. The capillary voltage and temperature were 42 V and 275°C, respectively. Selected Reaction Monitoring (SRM) analysis was performed using a TSQ Vantage (Thermo Fisher Scientific) operating in SRM mode. For SRM-mass spectrometry, the doubly charged precursor ions were monitored in Q1 with a resolution of 0.7 full width at half-maximum (FWHM) and singly charged y3 to n-1 ions for each peptide were monitored in Q3 at 0.7 FWHM. Each transition was monitored for 20 ms (dwell time) enabling a maximum duty cycle of 2.0 s. A digest of bovine serum albumin was analyzed every fifth run for signal intensity, retention time reproducibility, and peak width and shape to verify chromatography and instrument performance.

Targeted protein sequences for cathepsin B (cathB) were imported into Skyline (47) and converted into trypsin fragments. Methods were refined as described (6) with the following exception: all CID fragment y-ions (y3-yn-1) were monitored in all replicates. Three biological and three analytical replicates were analyzed and a Student's t-test was used to compare total peak areas.

The same P- and T-Myzus samples analyzed using 2-D DIGE were also subjected to a 1-D separation and analysis using LC-MS/MS. For each sample, 1 µg of total tryptic peptides were loaded onto a 5 cm C-18 PepMap trapping column with an EasyNanoLC 1000 (Thermo Fisher Scientific) system. The peptides were loaded with at a maximum pressure of 280 bar and were washed with 3 µL of 0.1% formic acid at the same pressure settings. Following sample cleanup, the peptides were separated by a 20cm hand-packed and pulled C-18 PepMap column where the voltage for ionization was applied at a liquid junction prior to the column. The gradient was a total of 230 minutes, beginning at 4% acetonitrile, 0.1% formic acid and ramping linearly to 35% acetonitirile, 0.1% formic acid in 185 minutes. The remaining time was used for a high organic and re-equilibration of the trapping and analytical columns. Nanospray ionization was achieved with a NanoFlex ion source (Thermo Fisher Scientific) operating at 1.9 kV and an ion transfer tube temperature of 275C.

MS/MS analysis was performed on a Q Exactive (Thermo Fisher Scientific) using software version 2.2 SP1. Profile MS1 spectra were obtained from an *m/z* of 400 to 1600 at 140k resolution with an AGC target of $3.0e^6$ charges and a maximum fill time of 50ms. The 20 most intense ions from each MS1 scan with

charge states of 2-7 were selected for fragmentation with preference given to those with an isotopic distribution matching that of average using the peptide match option “preferred”. Selected ions were isolated in a 1.6 Da window centered on the most intense isotope with an AGC target of $1e^5$ and a maximum fill time of 50 ms. Fragmentation was normalized to a collision energy of 27 for each selected ion. Targeted ions were then placed on the dynamic exclusion list for 15 sec to allow multiple fragmentations of each ion. Protein identification was done as described above for gel spots except that a different database was used. A custom database was constructed for the label-free analysis and included a total of 35,482 sequences, including the proteins from the former database as well as the predicted protein sequences from the pea aphid genome available on NCBI and common contaminant proteins. Additional changes for this analysis included one missed tryptic cleavage was allowed and methylthio was selected as a fixed modification (digests were treated with methyl methanethiosulfonate as described in (48). Spectral counting was performed using Scaffold (Proteome Software) and differentially expressed proteins were selected using a p-value <0.05 and the Fisher’s Exact test. Data are available via ProteomeXchange with identifier PXD004893.

CathB quantification in aphids at the transcript level. The levels of CathB transcripts were compared between P-Myzus and T-Myzus by qRT-PCR. Ten adult aphids were collected from both colonies and flash frozen for RNA isolation. Quantitative Reverse transcribed PCR (qRT-PCR) was used to measure cathB expression at the transcript level in T- and P-Myzus. Pools of ten whole aphids

were milled to a fine powder in a cryogrinder (Retsch) and total RNA was extracted by using the RNeasy mini kit (Qiagen), followed by cDNA synthesis using 1 ug of total RNA and the SuperScript III First Strand Synthesis kit (Invitrogen) with Oligo dT primers. qRT-PCR reactions were performed using 2 ng of cDNA and 10 uL of SYBR Green PCR Master Mix (Applied Biosystems). Gene specific primers were used to amplify CathB (5'- ACAAGCGACTACATGGAAGG- 3' and 5'-CCCAACACGATCCACAATTTC-3') and *B*-actin (5'-TCGTCTTGGATTCTGGTGATG-3' and 5'-GCAAGATCGAGACGAAGGATAG-3') *M. persicae* genes. Ct values of Cathepsin were normalized to the Ct values of the reference gene *B*-actin. Three biological replicates and three analytical replicates were performed for each gene.

Structural modeling. Structural modeling was performed using the Phyre2 Protein Fold Recognition Server in "normal" model (49). Structural models were visualized using Molsoft.

Co-immunolocalization of CathB and PLRV in *M.persicae* gut. Immunostaining assays were used to localize CathB in the aphid digestive system and to test the hypothesis that this protein co-localizes with PLRV in the aphid gut. The assays were performed with guts dissected from P- and T-Myzus fed on a PLRV-infected HNS detached leaves for an AAP of 48 hr, after which aphid guts were dissected and prepared as described (50). Briefly, guts dissected in 1x PBS were fixed in 4% paraformaldehyde and then permeated by adding 0.1% Triton X-100. After washing three times with 1x PBST (PBS+0.5% Tween 20), tissues were blocked for 1h in blocking buffer (1xPBST containing 1% bovine serum albumin) and

incubated overnight at 4°C with α-CathB human monoclonal antibody raised in rabbit (ABCAM ab125067). Midguts were washed three times with PBST and incubated for 1h at room temperature with donkey anti-rabbit IgG secondary antibody conjugated to Cy3 (Jackson Immunoresearch 711-165-152). For the co-localization, tissues were blocked again for 1 hr and then incubated with α-PLRV coat protein antibody at 4°C overnight. Tissues were then washed three times with PBST and incubated for 1 hr at room temperature with donkey α-rabbit IgG secondary antibody conjugated to Cy2 (Jackson Immunoresearch 711-225-152). Gut tissues were washed again for three times with PBST, mounted in 1x PBS buffer containing DAPI, covered and sealed for analysis under a confocal microscope. For each treatment, three guts were analyzed. Two different controls were performed: 1) healthy controls: aphids fed only on healthy plants and prepared with CathB and PLRV antibodies; 2) guts of aphids fed on PLRV-infected HNS leaves were prepared using only secondary antibodies (no primary antibodies).

Immunocapture RT-PCR. The potential interaction between the PLRV coat protein and *M. persicae* CathB was investigated by Immunocapture PCR as described (51) with the addition of the reverse transcription step. Briefly, the wells of a microtitre plate were coated in a coating buffer (0.05 M sodium carbonate pH 9.5) for 4 hr at room temperature with 4 µg of the following antibodies: α-CathB human monoclonal antibody raised in rabbit (ABCAM ab125067), α-CathB polyclonal antibody raised in rabbit (ABCAM ab92955), and as a positive control, α-PLRV coat protein (Agdia). Negative controls were performed with α-IgG

antibody and with no antibody. P- and T-Myzus aphids previously fed on PLRV-infected HNS detached leaves for 48 hr were homogenized in one volume of extraction buffer (50 mM Tris–HCl (pH 8.0), 5 mM EDTA, 2% polyvinylpyrolidone and 0.05% Tween-20) and centrifuged at 13,000 × g for 15 min at 4°C. The plate was washed with 1×PBS three times to remove unbound antibody, and 200 µl aliquots of aphid extract were added into the wells in three replicates. The plate was incubated overnight at 4°C to allow maximum capturing of the particles. The overnight-incubated plate was washed with 1× PBS to remove all unbound material. The bound particles were released by adding 30 µl/well of extraction buffer and the suspension was stored at 4°C until further use. A total of 5 µl of the extracted particles was used to synthesize cDNA by using Improm II reverse transcriptase, in a 15ul reverse transcription reaction. PCR amplification of the viral cDNA from the virions bound to the antibodies used in the capture was performed with 3ul of cDNA using specific primers to amplify a 660bp fragment of the PLRV coat protein (44).

CathB activity assays and inhibition of cysteine proteases using E-64. The activity of cathB in P-Myzus and in T-Myzus was compared in a fluorescence assay, by using the cathB activity assay kit from ABCAM (Cambridge, UK). A pool of adult aphids with the same weight (10 mg) was collected from both colonies, in three biological replicates. Aphids were ground in 200 µl of the lysis buffer provided in the kit in a cryogrinder with metal beads and incubated in ice for 10 min, followed by 5 min of centrifugation at 13,000 rpm and 4°C. Fifty µl of the aphid lysate was mixed with 50 µl of the CB reaction buffer provided in the kit in each well of a 96-

well plate. Two µl of the substrate Ac-RR-AFC was added to each sample/well and mixed by pippeting. The cysteine protease inhibitor E-64 (10 µM) was added to the negative controls. Samples were incubated at 37°C for 90 min and then read in a spectrofluorometer with 400 nm excitation and 505 nm emission filter. Relative Fluorescence Units (RFU) were compared between the two treatments.

In the second experiment, concentrated samples were prepared by grinding 35 mg of whole aphids in 200 µL of lysis buffer, as described above. Aphid lysate was then diluted to 10, 100 and 1000 fold to perform a sample dilution curve and then mixed with the CB reaction buffer provided in the kit (1:1) and the substrate Ac-RR-AFC, following incubation at 37°C and absorbance reading at a spectrofluorometer, as described above. A substrate dilution curve was also performed by diluting the substrate to 10, 100 and 1000 fold and keeping the sample concentration constant.

Cathepsin inhibition and PLRV transmission assays. Adult aphids were placed in dishes covered by a Parafilm® sandwich membrane containing an aphid diet prepared with balanced amounts of amino acids (45) for 48 hr. The adults were then removed, and first instar nymphs were transferred to a fresh diet containing the cysteine protease inhibitor E-64 (AG Scientific) at 0, 10, 30 or 50 µM, for a 48h AAP. Nymphs were transferred to PLRV-infected HNS detached leaves for a 24 hr AAP, after which ten nymphs per plant were then transferred to healthy potato cv. Red Maria seedlings for an IAP of 48 hr, with 20 replicates per treatment. Plants were treated with imidacloprid to kill the insects. After three weeks, PLRV systemic infection was confirmed by DAS-ELISA. These experiments were repeated three

times. The proportion of plants infected with PLRV was compared to non-infected plants using the Chi square statistic test.

ICP analysis of nutrients in aphids. Inductively Coupled Plasma Atomic Emission Spectrometer (ICP-AES) was used to quantify the amounts of macro and micronutrients in whole aphids (P- vs. T-Myzus), using three biological replicates per treatment. Aphid tissue was dried at 60°C for four days and then macerated using a mortar and pestle. Samples were weighed and then incubated overnight on 2.5 ml of 60:40 HNO3 and HClO4 mixture into a Pyrex glass tube to degrade organic matter. The mixture was then heated to 120°C for two hr and 0.25 ml of 40 ppm Yttrium added as an internal standard to compensate any drift during run in the ICP-AES. The temperature of the heating block was then raised to 190°C for 10 min and turned off. Samples were then cooled down and diluted to 20 ml, vortexed and transferred into auto sampler tubes to run in the ICP-AES Thermo iCAP 6500. The amount of each nutrient was converted from ppm to mg/g of sample based on the original weight of each sample. The normalized amount of each nutrient was compared between treatments by one-way ANOVA, and the means were compared using the Student's t-test.

RESULTS

Turnip-reared aphids are less efficient in transmitting PLRV as compared to physalis-reared aphids. When aphids were given an AAP of 24 hr, a significant difference in the proportion of plants infected with PLRV was observed (Fig. 2A).

P-Myzus infected 70% of the plants, while T-Myzus infected only 20% of the plants (Fig. 2A, p=0.035). When aphids were given a longer AAP of 48 hr, the transmission efficiency increased for both aphid lineages, but it was still significantly higher for P-Myzus (p=0.025). P-Myzus infected 80% of the plants with PLRV, while T-Myzus infected 33% of the plants with PLRV (Fig. 2A).

The host switch effect is transient. To test whether the host switch effect was transient, we transferred aphids from the physalis colony to turnip plants (PT-Myzus) and aphids from the turnip colony to physalis plants (TP-Myzus) for a three-day feeding period prior to allowing a 48 hr AAP on PLRV-infected HNS. The transmission efficiency was significantly different between the two groups of aphids (p=0.0235). The host switch caused the transmission phenotypes to be reversed. Aphids that were reared on physalis and then transferred to turnip for three days showed a dramatic decrease in their transmission efficiency, infecting 27.3% of the plants (Fig. 2B). Surprisingly, aphids from the turnip colony that were transferred to physalis for three days recovered their ability to efficiently transmit PLRV, infecting 72.7% of the plants (Fig. 2B).

The difference in transmission is not due to an inability to feed or colonize turnip. T-Myzus reproduced better compared to P-Myzus (Fig. 3). A ten-fold increase in the progeny number of T-Myzus as compared to P-Myzus ruled out the possibility that the difference in transmission was due to an inability to feed on or colonize turnip (Fig. 3A). Additionally, T-Myzus adults were double the mass of P-Myzus adults (Fig. 3B).

P-Myzus do not acquire more virus than T-Myzus. Virus acquisition by aphids is a specific process that requires crossing physical barriers in the aphid gut and salivary gland cells (52–54), a process which is mediated by receptor molecules in the cell membranes (2). Even in compatible aphid-virus species pairs, aphids do not acquire all viruses they ingest. Some virus particles might not cross the barriers presented by the aphid cells, others may be degraded. We tested whether the difference in PLRV transmission could be due to a difference in the aphids' ability to acquire the virus when they are reared on different host plants by comparing PLRV genomic RNA levels in T- and P-Myzus after gut clearing on artificial diets. Viral genomic RNA is unstable in the aphid gut (44, 55) and so the gut clearing step removes ingested virus that is not acquired and enables us to obtain an accurate measure of the virus that is acquired by the aphid. At both time points of 24 hr and 72 hr after virus acquisition there is a trend of higher number of PLRV copies in P-Myzus than in T-Myzus, although not significant (Fig. 4).

2-D Difference Gel Electrophoresis reveals that cathB is the major proteome change between T- and P-Myzus. The proteomes of T- and P-Myzus, were analyzed by quantitative 2-D DIfference Gel Electrophoresis (DIGE). A total of 1324 spots were visualized using DIGE on the Cy2 reference gel and matched across all gels in the experiment (Fig. 5). Since P- and T-Myzus have the same genetic background, it was expected that their proteome profile would have only a few differences in protein expression. Accordingly, of the 1324 total spots, only 15 spots were differentially expressed (Fig. 5) by at least a 2.5 fold change (p-value < 0.005) using a one-way ANOVA, all of them up-regulated in T-Myzus and all

located in one region of the 2-D gels (Fig. 5, spots in red). A Cy3-Cy5 dye swap on biological replicates (Fig. 5, inset) showed the pattern of protein spot up-regulation in the T-Myzus samples was highly reproducible and not an artifact of dye-labeling bias. The spot pattern is indicative of a protein with multiple isoforms varying by charge, with isoelectric points ranging from pH 4.2-5.5 (Fig. 5) and molecular mass from 45 to 35 kDa (Fig. 5). The proteins in these spots were identified as isoforms of the aphid cysteine protease, cathB, with at least eight matching peptides, using nLC-MS/MS analysis (Table 1). Peptides from lower abundant proteins (more than 20-fold lower) were identified in these spots, but this is common for gel spot analysis by LC-MS/MS. Other spots appearing as green or red on the image DIGE provided in Figure 5 did not represent statistically reproducible differences between T- and P-Myzus in the 2-D DIGE experiment. These spots may represent proteins with high biological variability requiring more replicates to show statistically different fold-changes, or analytical variation from 2-D gel, DIGE and/or protein extraction artifacts.

CathB in *M. persicae.* The draft genome of *M. persicae* is available from Aphidbase (http://www.aphidbase.com/node_94263/Myzus-DB/Downloads). Three cathB genes were identified from the available sequencing data (Supplementary Fig. S.1). CathB transcripts were 11-fold higher in T-Myzus compared to P-Myzus (Supplementary Fig. S.2), indicating that the difference in expression of this protein between T- and P-Myzus results from differences in transcript abundance.

Two genes mined from the sequencing data were full length, contigs 1 and 2, with isoelectric points (pIs) of 5.3 and 5.2 and molecular mass at 37 kDa, which corresponded to the protein spots identified on the 2-D gels. CathB contig 3 was truncated and predicted to be 14 kDa in size with an isoelectric point at 4.2. All cathB peptides isolated from the 2-D gels were either shared among all three contigs or between cathB contigs 1 and 3. To test whether cathB contig 2 was also up-regulated in T-Myzus, we used selected reaction monitoring mass spectrometry (SRM) targeting peptides that were shared and specific to cathepsin B contigs 1 and 2 (Fig. 6A and 6B, respectively). Peak areas from all peptides measured using SRM were higher in abundance in T-Myzus (Fig. 6), confirming that cathB protein expressed from both contigs 1 and 2 were up-regulated in T-Myzus. The structural topology of the three catalytic residues of *M. persicae* cathepsin B were conserved with human cathepsin B (Supplementary Fig. S.3), further verifying our annotation of these contigs as *M. persicae* cathB.

Other enzymes with predicted lysosomal function are upregulated in T-Myzus.

To expand on the 2-D DIGE analysis, we performed quantitative label-free proteomics comparing tryptic digests of P and T-Myzus. This analysis revealed that other enzymes with predicted lysosomal functions were up-regulated in T-Myzus, compared to P-Myzus, such as β-glucuronidase, peroxidasin, legumain-like, and aminopeptidase-N (Fig.7). CathB and cathB-16 were also identified as upregulated in T-Myzus in the label-free experiment (Fig.7). CathB-16 was the major proteome level change between P- and T-Myzus in this experiment, with a

29-fold higher expression in T-Myzus, suggesting that aphid feeding on turnip plants induces a higher expression of other enzymes in the aphid gut, including additional cysteine proteases.

CathB co-localized with PLRV in gut brush border membranes of T-Myzus but not P-Myzus. In guts dissected from P-Myzus, the majority of PLRV and cathB were in distinct subcellular compartments, with cathB localized to a particular organelle along the midgut epithelial cytoskeleton and the virus was diffused throughout the cell (Fig.8A-C). In contrast, nearly complete co-localization of Cathepsin B and PLRV was observed in T-Myzus guts on the cell membranes (Fig.8D-F). Although the assay is not quantitative, we observed that the amount of cathB in P-Myzus was lower than in T-Myzus, consistent with the proteome data, while the abundance of PLRV was clearly greater inside of P-Myzus gut epithelial cells, consistent with the quantitative PCR data. Although cathB and PLRV co-localize to the cell membranes in T-Myzus, no direct interaction was detected using immunocapture-RT-PCR with the cathB antibodies and PLRV-specific primers (not shown). The localization of the virus is markedly different in P and T-Myzus and co-localization with aphid cathepsin B only occurs in T-Myzus, predominately on the brush border membranes of the midgut lumen. Controls to demonstrate the specificity of each antibody with non-viruliferous aphids, antibody controls and controls for the emission and excitation of each cyanine dye in the double labeling experiments are shown in Figs S4-6, respectively.

CathB activity in aphids

The increased level cathB in T-Myzus, suggests that there may be a higher activity level in these aphids as compared to P-Myzus. To test this, we performed a cathB activity assay. The activity of cathepsin B was significantly higher in T-Myzus than in P-Myzus (Fig. 9). The cysteine protease inhibitor E-64 significantly reduced the activity of cathepsin B in both P- and T-Myzus (Fig. 9), enabling us to develop a functional assay using E-64 to test the effect of inhibiting cathB on virus transmission.

Cathepsin inhibition in T-Myzus recovered the efficient vectoring phenotype. Feeding aphids on a cysteine protease inhibitor, E-64, resulted in a contrasting phenotype (Fig. 10) in P- and T-Myzus. Upon treatment with E-64 in T-Myzus, transmission of PLRV increased with the dose of E-64 in the diet ($p<0.0001$), reaching a maximum of nearly 80% of the recipient plants becoming infected with 50 µM E-64 (Fig. 10A). In contrast, PLRV transmission by P-Myzus significantly decreased following treatment with E-64 (Fig. 10B, $p=0.0004$), although the level did not differ among the concentrations of E-64 (Fig. 10B). No aphid mortality was observed in the inhibitor treatments compared to the controls (data not shown). The contrasting phenotypes observed in these E-64 inhibitor studies suggests that when cathB is highly expressed and on the cell periphery, it negatively regulates virus acquisition and when cathB is contained in the lysosomes, its activity is required but not sufficient for proper virus acquisition into the gut.

Varying enzyme and substrate concentration showed evidence for a cathepsin inhibitor in P-Myzus. The initial rate of a catalyzed reaction is directly proportional to the enzyme concentration over a wide range (56). The presence of

inhibitors, that are, molecules combining with the enzyme or the substrate, can easily be detected by the failure of proportionality in experiments measuring enzymatic activity as a function of substrate or enzyme concentration. To test for the presence of an aphid endogenous cathB inhibitor in P or T-Myzus, we performed fluorescence cathB activity assays in T- and P-Myzus by varying enzyme and substrate concentrations (Fig. 11). Holding all other conditions constant, the initial rate of reaction increased faster with a rise of substrate concentration in T- compared to P-Myzus (Fig. 11A), suggesting the presence of an inhibitor in P-Myzus. Similarly, controlling for weight and varying the aphid lysate concentration showed that the P-Myzus sample lost proportionality (Fig. 11B), also providing evidence for a cathB inhibitor in the P-Myzus sample. It is not known whether the inhibitor is derived from the plant or the aphid.

There are higher levels of Ca^{2+} in T-Myzus.

Based on the fact that T-Myzus are larger and more fecund than P-Myzus conspecifics, the proteome data that showed a higher level of lysosomal enzymes in T-Myzus, the co-localization of PLRV with cathB in the membranes of T-Myzus midgut epithelial cells, and our ability to increase the virus transmission efficiency of T-Myzus with the cysteine protease inhibitor E-64, we hypothesized that there is an increase in Ca^{2+}-mediated lysosomal functions in T-Myzus and this directly or indirectly reduces acquisition and transmission of PLRV by T-Myzus. We measured the amounts of Ca and other nutrients in P- and T-Myzus. Accordingly, the amount of Ca is higher in T-Myzus than in P-Myzus (Fig. 12).

Discussion

The ability of *M. persicae* to transmit PLRV when reared on turnip was impaired, although T-Myzus presented a fitness advantage, producing more progeny and larger adults than P-Myzus. A significant growth reduction was reported by Rahbé and collaborators (57) when they fed *M. persicae* on transgenic oilseed rape plants (*Brassica napus*) expressing the cysteine protease inhibitor oryzacystatin, which also suggests that cathepsin activity in the aphid is required for proper growth and low levels of cathepsin impairs aphid growth. These observations are consistent with our studies showing that T-Myzus reproduced better and were larger than P-Myzus, which correlated with the higher levels of cathB in T-Myzus. Collectively, the data show that the reduction in virus transmission efficiency in T-Myzus is not due to an impairment of aphid growth when reared on turnip plants.

Two proteomics studies have examined the impact of the host on the aphid vector (58, 59). To our knowledge, our study is the first to look at the impact of a host switch on vectoring ability, i.e., the efficiency at which an insect vector transmits a virus, at the molecular level and revealed cathB expression may regulate virus transmission by aphids. Two experimental approaches were available for testing the role of cathB in virus transmission, RNA interference (RNAi), which has been shown to be functional in aphids (60–64), or the use of a chemical cathB inhibitor, E-64, a fungal-derived secondary metabolite that has specificity for thiol-proteases, including cathB but not cathepsin A or D (65). Since the cath gene family is highly expanded in the aphid (66), both approaches have a drawback of some degree of lack of specificity. CathB silencing constructs will

have off target effects and E-64 will inhibit other thiol proteases in the aphid gut. In our hands, silencing of cathB in the aphid using RNA interference was not amenable for conducting virus transmission experiments using different host plants due to the length of time required for transmission experiments (data not shown). Since the inhibition of cathB is irreversible by E-64, it was our method of choice and we probed the function of cathB in transmission using E-64. In addition, we sought to use a functional assay that interfered with enzyme function, which E-64 provided, not just reduce the level of the enzyme by lowering transcript levels. While RNAi is certainly a powerful approach for functional genomics and given the paucity of tools for functional genomics in aphids and other hemipterans, the vast array of fungal secondary metabolites with specificity for different proteins makes this approach a highly attractive alternative to RNAi for functional genomics in these insects. The proteome-guided cell biological analyses in our study show that feeding on turnip influences the expression of several predicted lysosomal enzymes in the aphid. It is possible and likely that other proteins, including other cysteine proteases, play a regulatory role, directly or indirectly, in virus acquisition. The functions of these additional proteins in virus transmission should be further investigated.

The cathB gene family is massively expanded in aphids, which is thought to be an adaptation for feeding on phloem sap (66). This raises an interesting question about the potential role of cathepsins in regulating the ability of an aphid to transmit viruses. Cathepsins are papain cysteine proteases under physiologic conditions (67) and are widely expressed in the aphid gut (68, 69), including in the

anterior aphid midgut (70). They function as both endopeptidases and as peptiddyldipeptases in the lysosome (67), cleaving substrate proteins as they enter into the organelle. It is notable that "putative cathepsin B-S" was the fifth most abundant transcript in the gut of the soybean aphid and "cathepsin B-16A" the tenth most abundant (71), indicating that cathB clearly has important functions in the aphid gut. In our study, cathB-16 was also identified as upregulated in T-Myzus by the label-free proteomics. Other cysteine proteases might have been inhibited by E-64 in the inhibition assays, as E-64 is a cysteine protease inhibitor, not specific to cathB. Therefore, a role for the gut luminal cysteine proteases such as cathL cannot be ruled out.

CathB and PLRV do not directly interact in immunocapture experiments and argues for an indirect association between presence of Cathepsin B on the gut surface, its upregulation, and reduced virus transmission. Aphids have multiple cathepsin genes that are tightly regulated in a tissue specific manner (66, 68, 69, 71). CathL and cathB are the major cathepsins in the aphid gut and cathB is one of the four most important detoxification enzymes (70, 72–74). CathB is directly involved in the hydrolysis of toxic proteins in the diet (79). Brassicaceae plants (including turnip) contain protease inhibitors that are overcome by upregulation of cathepsins in the gut (76). What is observed in our experiments is upregulation of and increase in activity of gut localized cathB in response to the host plant and an associated decrease in the transmission of PLRV.

A number of different hypotheses are possible to explain these observations. The first of these is that the effect is indirect resulting from the

presence of turnip proteins. Turnips contain lectins that bind to glycosylated proteins on the surface of the gut epithelium in herbivores (77). Such binding may sterically interfere with association of PLRV with its receptor. This scenario is unlikely because our data show that by inhibiting cathepsin activity using E-64, virus transmission efficiency can be restored. A second, and more plausible alternative is that the increase of cathepsin B in T-myzus is degrading turnip phloem proteins that facilitate virus entry. Previous work has shown evidence for plant proteins associating with purified luteovirids (6) and that phloem proteins may assist in virus transmission (78). E-64 may block cysteine protease mediated degradation of turnip proteins that facilitate virus entry in T-Myzus, rendering them poorer vectors than their P-Myzus conspecifics.

Other scenarios are also possible. Inhibition of cathepsins in the gut of P-Myzus could result in up-regulation of other proteolytic enzymes, which could have a similarly detrimental effect on the virus, and these other proteolytic enzymes may have slightly reduced the PLRV transmission efficiency. Reduced cysteine proteases in the gut have been reported to have detrimental effects on aphid gut cells (68). Damage to the gut cells could hamper virus acquisition into the aphid vector; however, no such effects were observed during our experiments. Our data suggest a biphasic effect of cathB in PLRV transmission by *M. persicae*, where some cathepsin activity is required inside a particular organelle in the midgut (possibly lysosomes), but very high levels of the enzyme on the brush border membranes inhibits virus transmission, as we observed in T-Myzus. Other non-mutually exclusive ideas include that different protein complexes are formed with

cathB with distinct functions and the expression of cathB in the gut cell membranes of T-Myzus also relocalizes a receptor critical for virus acquisition; the pH of the gut lumens are different and this change protease activity which changes vector competency; or that the proteolytic specificities of the enzyme are not the same in T- and P-Myzus due to their distinct subcellular localizations. Finally, the larger amount of calcium in T-Myzus as compared to P-Myzus may cause an increase in lysosomal activities in the aphid midgut, a hypothesis that is also consistent with the label-free proteomics data showing an increase in the expression of other enzymes with predicted lysosomal function in T-Myzus. As lysosomes are the hub of cellular homeostasis (79) ,an increase in lysosomal exocytosis or other lysosomal functions may impair intracellular trafficking, egress, or uptake of the virus or one of its receptors.

The generalist aphid *M. persicae* feeds on a large array of host plants, from more than 40 plant families (80), which requires a broad adaptive capacity and phenotypic plasticity. This aphid is one of the most important vectors of PLRV (23, 81, 82). It has been shown in the past that the host plant on which this aphid is reared on influences its vector competence (83–85). For example, a difference in virus transmission was observed for *Zucchini Yellow Mosaic Virus* (ZYMV), a virus that is nonpersistently transmitted by *M. persicae,* when aphids were reared on two plants that are both hosts of *M. persicae* but non-hosts of ZYMV, the Brassica mustard (*Brassica juncea*) and the Malvaceous okra (*Abelmoschus esculentus*) (85). Overall, aphids reared on mustard plants had higher transmission rates of ZYMV to host recipient plants (*Cucurbita pepo*) than okra-reared aphids.

Interestingly, the host switch had a complex effect, giving intermediate transmission rates when mustard-reared aphids were given a preacquisition time of 24 hours on okra plants, showing that the host switch effect was transient and in a continuum. In a previous study with PLRV, 75% of the nymphs born on physalis transmitted PLRV to recipient plants, while only 49% of the nymphs born on rape (*Brassica rape)* transmitted the virus (84). Similarly to our study, these authors used a host plant of PLRV, physalis, and a nonhost of PLRV, the Brassica rape. A similar effect was observed in another study, when *M. persicae* reared on rape were less efficient in transmitting *Beet Yellows Virus* (BYV) than aphids reared on beet, a host of the virus (83). Again, in this case, rape is a nonhost of BYV. Therefore, it seems that rearing the aphids on a host plant of the virus increases the efficiency of virus transmission by this generalist aphid, which may have implications for virus epidemiology. Our data add to this body of knowledge and for the first time provides important insights into how the host plant is impacting vectoring ability at the molecular level with the support of proteomics, biochemical and imaging data.

References

1. **Tamborindeguy C**, **Bereman MS**, **DeBlasio S**, **Igwe D**, **Smith DM**, **White F**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2013. Genomic and proteomic analysis of *Schizaphis graminum* reveals cyclophilin proteins are involved in the transmission of *Cereal yellow dwarf virus*. PloS One **8**:e71620.

2. **Linz LB**, **Liu S**, **Chougule NP**, **Bonning BC**. 2015. In vitro evidence supports membrane alanyl aminopeptidase n as a receptor for a plant virus in the pea aphid vector. J Virol JVI.01479-15.

3. **Gray S**, **Cilia M**, **Ghanim M**. 2014. Circulative, “nonpropagative” virus transmission: an orchestra of virus-, insect-, and plant-derived instruments. Adv Virus Res **89**:141–199.

4. **Whitfield AE**, **Falk BW**, **Rotenberg D**. 2015. Insect vector-mediated transmission of plant viruses. Virology **479**–**480**:278–289.

5. **Tamada T**, **Harrison BD**. 1981. Quantitative studies on the uptake and retention of *Potato leafroll virus* by aphids in laboratory and field conditions. Ann Appl Biol **98**:261–276.

6. **Cilia M**, **Peter KA**, **Bereman MS**, **Howe K**, **Fish T**, **Smith D**, **Gildow F**, **MacCoss MJ**, **Thannhauser TW**, **Gray SM**. 2012. Discovery and targeted LC-MS/MS of purified polerovirus reveals differences in the virus-host interactome associated with altered aphid transmission. PloS One **7**:e48177.

7. **Brault V**, **Périgon S**, **Reinbold C**, **Erdinger M**, **Scheidecker D**, **Herrbach E**, **Richards K**, **Ziegler-Graff V**. 2005. The polerovirus minor capsid protein determines vector specificity and intestinal tropism in the aphid. J Virol **79**:9685–9693.

8. **Reinbold C**, **Herrbach E**, **Brault V**. 2003. Posterior midgut and hindgut are both sites of acquisition of Cucurbit aphid-borne yellows virus in *Myzus persicae* and *Aphis gossypii*. J Gen Virol **84**:3473–3484.

9. **Gildow FE**, **Reavy B**, **Mayo MA**, **Duncan GH**, **Woodford JA**, **Lamb JW**, **Hay RT**. 2000. Aphid acquisition and cellular transport of *Potato leafroll virus*-like particles lacking P5 Readthrough protein. Phytopathology **90**:1153–1161.

10. **Garret A**, **Kerlan C**, **Thomas D**. 1996. Ultrastructural study of acquisition and retention of potato leafroll luteovirus in the alimentary canal of its aphid vector, *Myzus persicae* Sulz. Arch Virol **141**:1279–1292.

11. **Garret A**, **Kerlan C**, **Thomas D**. 1991. Visualization of a luteovirus in the vector aphid's body by two gold immunolabelling techniques: a comparative study. J Virol Methods **35**:93–107.

12. **Rouzé-Jouan J**, **Terradot L**, **Pasquer F**, **Tanguy S**, **Giblot Ducray-Bourdin D D**. 2001. The passage of *Potato leafroll virus* through *Myzus persicae* gut membrane regulates transmission efficiency. J Gen Virol **82**:17–23.

13. **Garret A**, **Kerlan C**, **Thomas D**. 1993. The intestine is a site of passage for *Potato leafroll virus* from the gut lumen into the haemocoel in the aphid vector, *Myzus persicae* Sulz. Arch Virol **131**:377–392.

14. **Gildow FE**. 1987. Virus—membrane interactions involved in circulative transmission of luteoviruses by aphids, p. 93–120. *In* Harris, KF (ed.), Current Topics in Vector Research. Springer New York.

15. **Liu S**, **Sivakumar S**, **Sparks WO**, **Miller WA**, **Bonning BC**. 2010. A peptide that binds the pea aphid gut impedes entry of *Pea enation mosaic virus* into the aphid hemocoel. Virology **401**:107–116.

16. **Shepardson S**, **Esau K**, **McCrum R**. 1980. Ultrastructure of potato leaf phloem infected with *Potato leafroll virus*. Virology **105**:379–392.

17. **Peiffer ML**, **Gildow FE**, **Gray SM**. 1997. Two distinct mechanisms regulate luteovirus transmission efficiency and specificity at the aphid salivary gland. J Gen Virol **78 (Pt 3)**:495–503.

18. **Gildow FE**. 1993. Evidence for receptor-mediated endocytosis regulating luteovirus acquisition by aphids. Phytopathol USA.

19. **Ingwell LL**, **Eigenbrode SD**, **Bosque-Pérez NA**. 2012. Plant viruses alter insect behavior to enhance their spread. Sci Rep **2**:578.

20. **Mayer RT**, **Inbar M**, **McKenzie CL**, **Shatters R**, **Borowicz V**, **Albrecht U**, **Powell CA**, **Doostdar H**. 2002. Multitrophic interactions of the silverleaf whitefly, host plants, competing herbivores, and phytopathogens. Arch Insect Biochem Physiol **51**:151–169.

21. **Gildow FE**. 1980. Increased production of alatae by aphids reared on oats infected with *Barley yellow dwarf virus*. Ann Entomol Soc Am **73**:343–347.

22. **Castle SJ**, **Mowry TM**, **Berger PH**. 1998. Differential settling by *Myzus persicae* (Homoptera: Aphididae) on various virus infected host plants. Ann Entomol Soc Am **91**:661–667.

23. **Eigenbrode SD**, **Ding H**, **Shiel P**, **Berger PH**. 2002. Volatiles from potato plants infected with *Potato leafroll virus* attract and arrest the virus vector, *Myzus persicae* (Homoptera: Aphididae). Proc Biol Sci **269**:455–460.

24. **Hodge S**, **Powell G**. 2008. Do plant viruses facilitate their aphid vectors by inducing symptoms that alter behavior and performance? Environ Entomol **37**:1573–1581.

25. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. Proc Natl Acad Sci U S A **107**:3600–3605.

26. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Effects of *Cucumber mosaic virus* infection on vector and non-vector herbivores of squash. Commun Integr Biol **3**:579–582.

27. **Mauck K**, **Bosque-Pérez NA**, **Eigenbrode SD**, **De Moraes CM**, **Mescher MC**. 2012. Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses. Funct Ecol **26**:1162–1175.

28. **McMenemy LS**, **Hartley SE**, **MacFarlane SA**, **Karley AJ**, **Shepherd T**, **Johnson SN**. 2012. Raspberry viruses manipulate the behaviour of their insect vectors. Entomol Exp Appl **144**:56–68.

29. **Miller JW**, **Coon BF**. 1964. The effect of Barley yellow dwarf virus on the biology of its vector the English grain aphid, *Macrosiphum granarium*. J Econ Entomol **57**:970–974.

30. **Castle SJ**, **Berger PH**. 1993. Rates of growth and increase of *Myzus persicae* on virus-infected potatoes according to type of virus-vector relationship. Entomol Exp Appl **69**:51–60.

31. **Blua MJ**, **Perring TM**, **Madore MA**. 1994. Plant virus-induced changes in aphid population development and temporal fluctuations in plant nutrients. J Chem Ecol **20**:691–707.

32. **Luan J-B**, **Li J-M**, **Varela N**, **Wang Y-L**, **Li F-F**, **Bao Y-Y**, **Zhang C-X**, **Liu S-S**, **Wang X-W**. 2011. Global analysis of the transcriptional response of whitefly to *Tomato yellow leaf curl China virus* reveals the relationship of coevolved adaptations. J Virol **85**:3330–3340.

33. **Luan J-B**, **Yao D-M**, **Zhang T**, **Walling LL**, **Yang M**, **Wang Y-J**, **Liu S-S**. 2013. Suppression of terpenoid synthesis in plants by a virus promotes its mutualism with vectors. Ecol Lett **16**:390–398.

34. **Matsuura S**, **Hoshino S**. 2009. Effect of tomato yellow leaf curl disease on reproduction of *Bemisia tabaci* Q biotype (Hemiptera: Aleyrodidae) on tomato [*Lycopersicon esculentum*] plants. Appl Entomol Zool Jpn.

35. **Maris PC**, **Joosten NN**, **Goldbach RW**, **Peters D**. 2004. To*mato spotted wilt virus* infection improves host suitability for its vector *Frankliniella occidentalis*. Phytopathology **94**:706–711.

36. **Belliure B**, **Janssen A**, **Maris PC**, **Peters D**, **Sabelis MW**. 2005. Herbivore arthropods benefit from vectoring plant viruses. Ecol Lett **8**:70–79.

37. **Nanayakkara UN**, **Giguère MA**, **Pelletier Y**. 2012. Population growth of *Myzus persicae* on potato plants infected with different strains and variants of potato virus Y. Am J Potato Res **90**:297–300.

38. **Wosula EN**, **Davis JA**, **Clark CA**. 2013. Population dynamics of three aphid species (Hemiptera: Aphididae) on four Ipomoea spp. infected or noninfected with sweetpotato potyviruses. J Econ Entomol **106**:1566–1573.

39. **Ellsbury MM**, **Pratt RG**, **Knight WE**. 1985. Effects of single and combined infection of arrowleaf clover with Bean yellow mosaic virus and a phytophthora sp. on reproduction and colonization by pea aphids (Homoptera: Aphididae). Environ Entomol **14**:356–359.

40. **Inoue T**, **Sakurai T**. 2006. Infection of *Tomato spotted wilt virus* (TSWV) shortens the life span of thelytokous *Thrips tabaci* (Thysanoptera: Thripidae). Appl Entomol Zool **41**:239–246.

41. **Donaldson JR**, **Gratton C**. 2007. Antagonistic effects of soybean viruses on soybean aphid performance. Environ Entomol **36**:918–925.

42. **Power A**, **Seaman A**, **Gray S**. 1991. Aphid transmission of *Barley yellow dwarf virus* - inoculation access periods and epidemiologic implications. Phytopathology **81**:545–548.

43. **Sadeghi E**, **Dedryver CA**, **Gauthier JP**. 1997. Role of acquisition and inoculation time in the expression of clonal variation for BYDV-PAV transmission in the aphid species *Rhopalosiphum padi*. Plant Pathol **46**:502–508.

44. **Peter KA**, **Liang D**, **Palukaitis P**, **Gray SM**. 2008. Small deletions in the *Potato leafroll virus* readthrough protein affect particle morphology, aphid transmission, virus movement and accumulation. J Gen Virol **89**:2037–2045.

45. **Kim JH**, **Jander G**. 2007. *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. Plant J Cell Mol Biol **49**:1008–1019.

46. **Cilia M**, **Fish T**, **Yang X**, **McLaughlin M**, **Thannhauser TW**, **Gray S**. 2009. A comparison of protein extraction methods suitable for gel-based proteomic studies of aphid proteins. J Biomol Tech JBT **20**:201–215.

47. **MacLean B**, **Tomazela DM**, **Shulman N**, **Chambers M**, **Finney GL**, **Frewen B**, **Kern R**, **Tabb DL**, **Liebler DC**, **MacCoss MJ**. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinforma Oxf Engl **26**:966–968.

48. **DeBlasio SL**, **Johnson R**, **Mahoney J**, **Karasev A**, **Gray SM**, **MacCoss MJ**, **Cilia M**. 2015. Insights into the polerovirus-plant interactome revealed by coimmunoprecipitation and mass spectrometry. Mol Plant-Microbe Interact MPMI **28**:467–481.

49. **Kelley LA**, **Sternberg MJE**. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc **4**:363–371.

50. **Pakkianathan BC**, **Kontsedalov S**, **Lebedev G**, **Mahadav A**, **Zeidan M**, **Czosnek H**, **Ghanim M**. 2015. Replication of *Tomato yellow leaf curl virus* in its whitefly vector, *Bemisia tabaci*. J Virol **89**:9791–9803.

51. **Gottlieb Y**, **Zchori-Fein E**, **Mozes-Daube N**, **Kontsedalov S**, **Skaljac M**, **Brumin M**, **Sobol I**, **Czosnek H**, **Vavre F**, **Fleury F**, **Ghanim M**. 2010. The transmission efficiency of *Tomato yellow leaf curl virus* by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. J Virol **84**:9310–9317.

52. **Li C**, **Cox-Foster D**, **Gray SM**, **Gildow F**. 2001. Vector specificity of *Barley yellow dwarf virus* (BYDV) transmission: identification of potential cellular receptors binding BYDV-MAV in the aphid, *Sitobion avenae*. Virology **286**:125–133.

53. **Gray S**, **Gildow FE**. 2003. Luteovirus-aphid interactions. Annu Rev Phytopathol **41**:539–566.

54. **Gildow F**, **Rochow W**. 1980. Role of accessory salivary-glands in aphid transmission of *Barley yellow dwarf virus*. Virology **104**:97–108.

55. **van den Heuvel JF**, **Bruyère A**, **Hogenhout SA**, **Ziegler-Graff V**, **Brault V**, **Verbeek M**, **van der Wilk F**, **Richards K**. 1997. The N-terminal region of the luteovirus readthrough domain determines virus binding to *Buchnera* GroEL and is essential for virus persistence in the aphid. J Virol **71**:7258–7265.

56. **Mitchell PH**. 1946. A textbook of Biochemistry. 1st. Ed. Mcgraw-Hill Book Company, Inc.; New York.

57. **Rahbé Y**, **Deraison C**, **Bonadé-Bottino M**, **Girard C**, **Nardon C**, **Jouanin L**. 2003. Effects of the cysteine protease inhibitor oryzacystatin (OC-I) on different aphids and reduced performance of *Myzus persicae* on OC-I expressing transgenic oilseed rape. Plant Sci **164**:441–450.

58. **Pinheiro P**, **Bereman MS**, **Burd J**, **Pals M**, **Armstrong S**, **Howe KJ**, **Thannhauser TW**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2014. Evidence of the biochemical basis of host virulence in the greenbug aphid, *Schizaphis graminum* (Homoptera: Aphididae). J Proteome Res **13**:2094–2108.

59. **Francis F**, **Gerkens P**, **Harmel N**, **Mazzucchelli G**, **De Pauw E**, **Haubruge E**. 2006. Proteomics in *Myzus persicae*: effect of aphid host plant switch. Insect Biochem Mol Biol **36**:219–227.

60. **Mutti NS**, **Park Y**, **Reese JC**, **Reeck GR**. 2006. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. J Insect Sci **6**.

61. **Pitino M**, **Coleman AD**, **Maffei ME**, **Ridout CJ**, **Hogenhout SA**. 2011. Silencing of aphid genes by dsRNA feeding from plants. PLoS ONE **6**.

62. **Jaubert-Possamai S**, **Le Trionnaire G**, **Bonhomme J**, **Christophides GK**, **Rispe C**, **Tagu D**. 2007. Gene knockdown by RNAi in the pea aphid *Acyrthosiphon pisum*. BMC Biotechnol **7**:63.

63. **Shakesby AJ**, **Wallace IS**, **Isaacs HV**, **Pritchard J**, **Roberts DM**, **Douglas AE**. 2009. A water-specific aquaporin involved in aphid osmoregulation. Insect Biochem Mol Biol **39**:1–10.

64. **Whyard S**, **Singh AD**, **Wong S**. 2009. Ingested double-stranded RNAs can act as species-specific insecticides. Insect Biochem Mol Biol **39**:824–832.

65. **Hanada K**, **Tamai M**, **Yamagishi M**, **Ohmura S**, **Sawada J**, **Tanaka I**. 1978. Isolation and characterization of E-64, a new thiol protease inhibitor. Agric Biol Chem **42**:523–528.

66. **Rispe C**, **Kutsukake M**, **Doublet V**, **Hudaverdian S**, **Legeai F**, **Simon J-C**, **Tagu D**, **Fukatsu T**. 2008. Large gene family expansion and variable selective pressures for cathepsin B in aphids. Mol Biol Evol **25**:5–17.

67. **Mort JS**, **Buttle DJ**. 1997. Cathepsin B. Int J Biochem Cell Biol **29**:715–720.

68. **Sapountzis P**, **Duport G**, **Balmand S**, **Gaget K**, **Jaubert-Possamai S**, **Febvay G**, **Charles H**, **Rahbé Y**, **Colella S**, **Calevro F**. 2014. New insight into the RNA interference response against cathepsin-L gene in the pea aphid, *Acyrthosiphon pisum*: molting or gut phenotypes specifically induced by injection or feeding treatments. Insect Biochem Mol Biol **51**:20–32.

69. **Deraison C**, **Darboux I**, **Duportets L**, **Gorojankina T**, **Rahbé Y**, **Jouanin L**. 2004. Cloning and characterization of a gut-specific cathepsin L from the aphid *Aphis gossypii*. Insect Mol Biol **13**:165–177.

70. **Cristofoletti PT**, **Ribeiro AF**, **Deraison C**, **Rahbé Y**, **Terra WR**. 2003. Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrthosiphon pisum*. J Insect Physiol **49**:11–24.

71. **Liu S**, **Chougule NP**, **Vijayendran D**, **Bonning BC**. 2012. Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. PloS One **7**:e45161.

72. **Silva AX**, **Bacigalupe LD**, **Luna-Rudloff M**, **Figueroa CC**. 2012. Insecticide resistance mechanisms in the green peach aphid *Myzus persicae* (Hemiptera: Aphididae) II: costs and benefits. PloS One **7**:e36810.

73. **Silva AX**, **Jander G**, **Samaniego H**, **Ramsey JS**, **Figueroa CC**. 2012. Insecticide resistance mechanisms in the green peach aphid *Myzus persicae* (Hemiptera: Aphididae) I: a transcriptomic survey. PLOS ONE **7**:e36366.

74. **Pyati P**, **Bandani AR**, **Fitches E**, **Gatehouse JA**. 2011. Protein digestion in cereal aphids (*Sitobion avenae*) as a target for plant defence by endogenous proteinase inhibitors. J Insect Physiol **57**:881–891.

75. **Koo YD**, **Ahn J-E**, **Salzman RA**, **Moon J**, **Chi YH**, **Yun D-J**, **Lee SY**, **Koiwa H**, **Zhu-Salzman K**. 2008. Functional expression of an insect cathepsin B-like counter-defence protein. Insect Mol Biol **17**:235–245.

76. **Jamal F**, **Pandey PK**, **Singh D**, **Khan MY**. 2013. Serine protease inhibitors in plants: nature's arsenal crafted for insect predators. Phytochem Rev **12**:1–34.

77. **Louda S**, **Mole S**. 1991. Glucosinolates: Chemistry and ecology, pp. 123–164, in G. A. Rosenthal and M. R. Berenbaum (eds.). Herbivores: Their Interactions with Secondary Plant Metabolites.

78. **Bencharki B**, **Boissinot S**, **Revollon S**, **Ziegler-Graff V**, **Erdinger M**, **Wiss L**, **Dinant S**, **Renard D**, **Beuve M**, **Lemaitre-Guillier C**, **Brault V**. 2010. Phloem protein partners of Cucurbit aphid borne yellows virus: possible involvement of phloem proteins in virus transmission by aphids. Mol Plant-Microbe Interact MPMI **23**:799–810.

79. **Medina DL**, **Di Paola S**, **Peluso I**, **Armani A**, **De Stefani D**, **Venditti R**, **Montefusco S**, **Scotto-Rosato A**, **Prezioso C**, **Forrester A**, **Settembre C**, **Wang W**, **Gao Q**, **Xu H**, **Sandri M**, **Rizzuto R**, **De Matteis MA**, **Ballabio A**.

2015. Lysosomal calcium signalling regulates autophagy through calcineurin and TFEB. Nat Cell Biol **17**:288–299.

80. **Blackman R**, **Eastop V**. 2000. Wiley: Aphids on the world's crops: an identification and information guide, 2nd Edition - R. L. Blackman, V. F. Eastop.

81. **van den Heuvel JFJM**, **Verbeek M**, **van der Wilk F**. 1994. Endosymbiotic bacteria associated with circulative transmission of *Potato leafroll virus* by *Myzus persicae*. J Gen Virol **75**:2559–2565.

82. **Ngumbi E**, **Eigenbrode SD**, **Bosque-Perez NA**, **Ding H**, **Rodriguez A**. 2007. *Myzus persicae* is arrested more by blends than by individual compounds elevated in headspace of PLRV-Infected potato. J Chem Ecol **33**:1733–1747.

83. **Gladders DW**, **Peters D**. 1986. The effect of previous host plant on the fecundity of *Myzus persicae* and its ability to transmit *Beet yellows virus*. Ann Appl Biol **109**:499–507.

84. **Mac Kinnon JP**. 1963. Some factors that affect the aphid transmission of two viruses that persist in the vector. Virology **20**:281–287.

85. **Symmes EJ**, **Perring TM**. 2007. Intraspecific variation in *Zucchini yellow mosaic virus* transmission by *Myzus persicae* and the impact of aphid host plant. J Econ Entomol **100**:1764–1772.

Funding Information

National Science Foundation Grant 1354309 to Michelle Cilia

National Institute of Child Health and Human Development Grant 5T15HD072999-04 to Michelle Cilia

Embrapa Student Fellowship to Patricia V. Pinheiro

Acknowledgements

We gratefully acknowledge members of the Cilia lab, Yi Xu (Cornell University) and Jim Bruce (University of Washington) for critical feedback on experiments and the manuscript, Jason Ingram (USDA ARS) for support with pesticide applications, Dawn Smith (Cornell, Retired) for assistance with aphid and plant husbandry, Kevin Howe (USDA ARS) for uploading the mass spectrometry data to ProteomeXChange, and Eric Craft and Shree Giri (USDA ARS) for performing the ICP analyses.

Figure Legends

Figure 1. Experimental design of transmission assays. Aphids reared on the two host plants, turnip and physalis, were transferred to PLRV-infected hairy nightshade (HNS) plants for an acquisition access period (AAP) of 24h or 48h, and then transferred to healthy potato seedlings for a virus inoculation for a 48h inoculation access period. Three weeks later, potato plants were checked for PLRV infection by DAS-ELISA.

Figure 2. PLRV transmission efficiency is regulated by vector-plant interactions. A) Proportion of potato plants cv. NY129 infected with *Potato leafroll virus* (PLRV) by *Myzus persicae* reared on physalis plants (P-Myzus) or on turnip plants (T-Myzus) after an acquisition access period (AAP) of 24 hours (n=10 plants, 5 insects/plant, p-value=0.035) or an AAP of 48 hours (n=10 plants, 10 insects/plant, p-value=0.025). B) Proportion of potato plants infected with PLRV by P-Myzus fed on turnip for 3 days (PT-Myzus) or T-Myzus fed on physalis for 3 days (TP-Myzus). PT- and TP-Myzus were fed on a source of PLRV for an AAP of 24 hours and then transferred to potato plants for an inoculation access period of 48 hours (n=15 plants, 5 insects/plant, p=0.0235).

Figure 3. *Myzus persicae* reared on turnip has fitness advantages. A) Average number of progeny of *M. persicae* reared on physalis plants (P-Myzus) or on turnip plants (T-Myzus) for 15 days (n=3 plants, 10 adults/plant, $p<0.001$). B) Average weight of adults of P-Myzus and T-Myzus (n=30, $p<0.001$).

Figure 4. Number of copies of *Potato leafroll virus* (PLRV) and standard error bars in *Myzus persicae* reared on physalis (P-Myzus) and turnip plants (T-Myzus), quantified by digital drop PCR.

Figure 5. 2-D Difference In Gel Electrophoresis shows minimal proteome level differences between *Myzus persicae* reared on turnip (T-Myzus) and physalis (P-Myzus). In this image, red spots are specific to T-Myzus, green spots specific to P-Myzus. Yellow spots are proteins expressed equally in both aphid colonies. Red

spot train was identified as cathepsin B (cathB) using mass spectrometry. Inset shows cathB spots in a dye-swap experiment.

Figure 6. Selected reaction monitoring reveals at least two cath B isoforms are up-regulated in *Myzus persicae* reared on turnip (T-Myzus) compared to physalis (P-Myzus): A) CathB isoform 1, B) CathB isoform 2. Two peptides abbreviated as DQG and DYY, are shared between the two cathepsin contigs. The four remaining peptides are unique to each cathB contig. The full tryptic peptide sequences corresponding to the three amino acid abbreviations can be found in Supplementary Table S.2.

Figure 7. Proteins upregulated in aphids reared on turnip (T-Myzus), compared to aphids reared on physalis (P-Myzus), measured by Mass Spectrometry.

Figure 8. Co-immunolocalization of CathB and PLRV in P-Myzus (A-C) and T-Myzus (D-F) alimentary canals. C is one focal plane in higher magnification of the inset in B, and F is a higher magnification of the inset in E. Blue in all panels is DAPI staining of the nuclei. Red is immunostaining of PLRV using specific primary antibody and secondary antibody conjugated to Cy2. Green is immunostaining of cathepsin B using specific primary antibody and secondary antibody conjugated to Cy3. Co-localization of PLRV and cathepsin B appears in yellow. ag: anterior midgut (stomach); pg: posterior midgut; hg: hindgut; e: embryo. Note that localization of PLRV and PLRV-CathB colocalization appears in some or all portions of the posterior midgut only.

Figure 9. CathB enzymatic activity assays on turnip-reared (T-Myzus) and physalis-reared (P-Myzus) *Myzus persicae*, before and after treatment with the cysteine protease inhibitor E-64.

Figure 10. Cathepsin inhibition modulates PLRV transmission by *Myzus persicae*. Proportion of plants infected with PLRV by aphids reared on turnip (T-Myzus) A) and on physalis (P-Myzus) B) after feeding on different concentrations of the cysteine protease inhibitor E-64.

Figure 11. Cathepsin enzymatic activity in aphids reared on turnip (T-Myzus) and physalis (P-Myzus) measured by a fluorescence activity assay and by varying substrate A) and sample lysate concentration B).

Figure 12. Calcium levels in turnip-reared (T-Myzus) and physalis-reared (P-Myzus) *Myzus persicae*, quantified by Inductively Coupled Plasma Atomic Emission Spectrometer (ICP-AES).

Table 1. Cathepsin B peptides identified using nLC-MS/MS analysis of in gel tryptic digests from protein isoform train in T-Myzus samples.

Myzus persicae Contig[a]	Peptide Sequence[b]	PSM E-value[c]
Contig 1	GLVTGGDYK	0.014
Contig 1	KGLVTGGDYK	8.3E-05
Contig 1	SGEGCEPYR	1.3E-04
Contig 1	SEDADYDNTYIPR	0.01
Contig 1	WRHCSTIGR	0.02
Contig 1	DYYYLTYGSIQK	0.005
Contig 1	MCYGDQDLDFDEDHR	4.6E-08
Contig 1	VPPCPNDDQGNNTCAGK PMESNHR	0.006

a: All peptides from the in gel digests matched to *M. persicae* contig 1, although other cathepsin B peptides from other aphid species present in the data base were also matched but not shown.

b: Peptide sequence, modified amino acids are underlined, oxidized methionine

c: Peptide Spectral Match (PSM) expect value reported by Mascot. Only peptides with E-values lower than 0.05 are reported here. All reported peptides were identified multiple times from replicate in gel digests.

Figure 1.

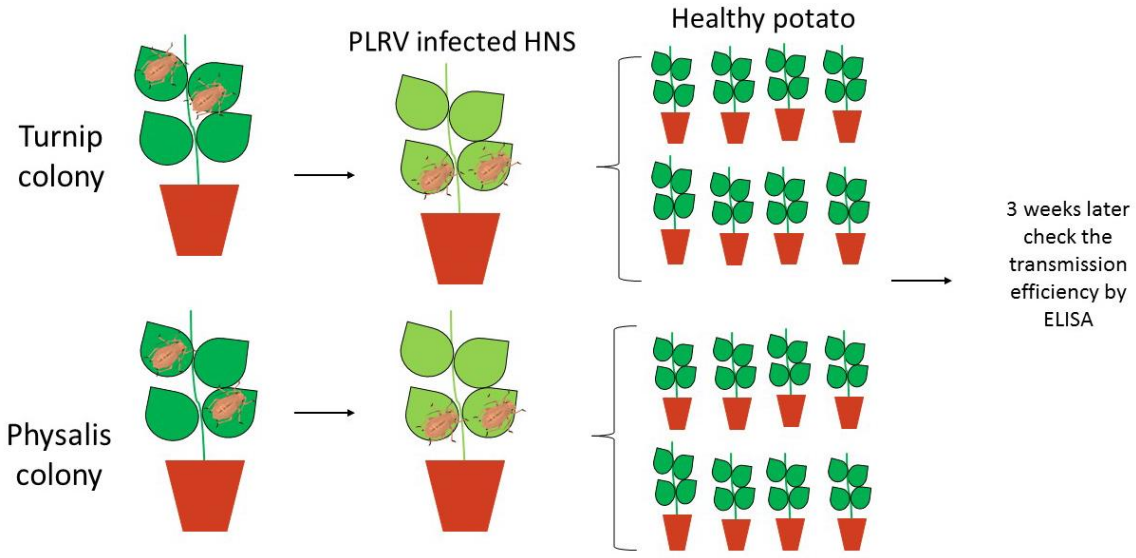

Turnip colony
Physalis colony
PLRV infected HNS
Healthy potato
3 weeks later check the transmission efficiency by ELISA

Figure 2.

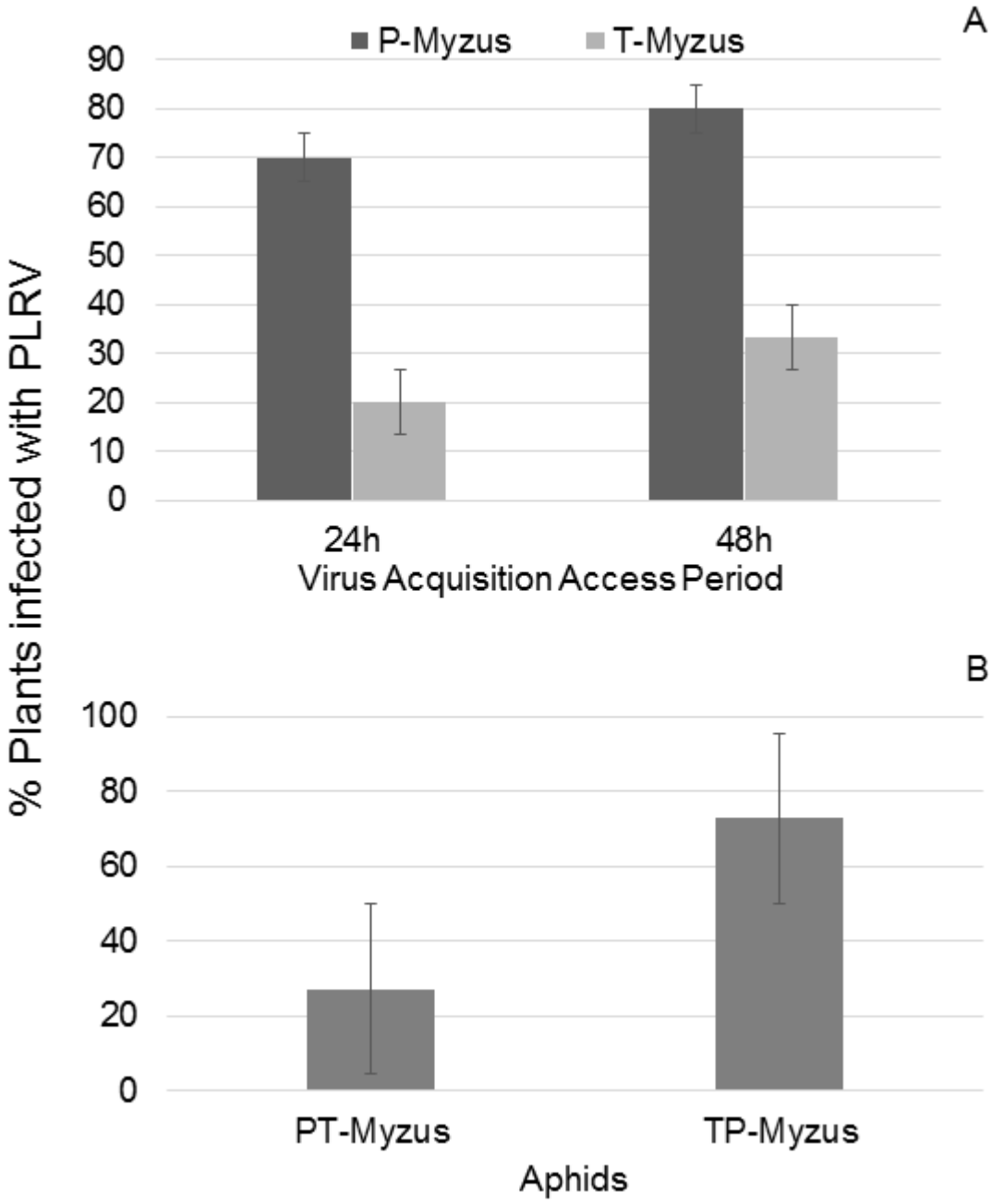

A
P-Myzus
T-Myzus
90
80
70
60
50
40
30
20
10
0
% Plants infected with PLRV
24h
48h
Virus Acquisition Access Period
B
100
80
60
40
20
0
PT-Myzus
TP-Myzus
Aphids

Figure 3.

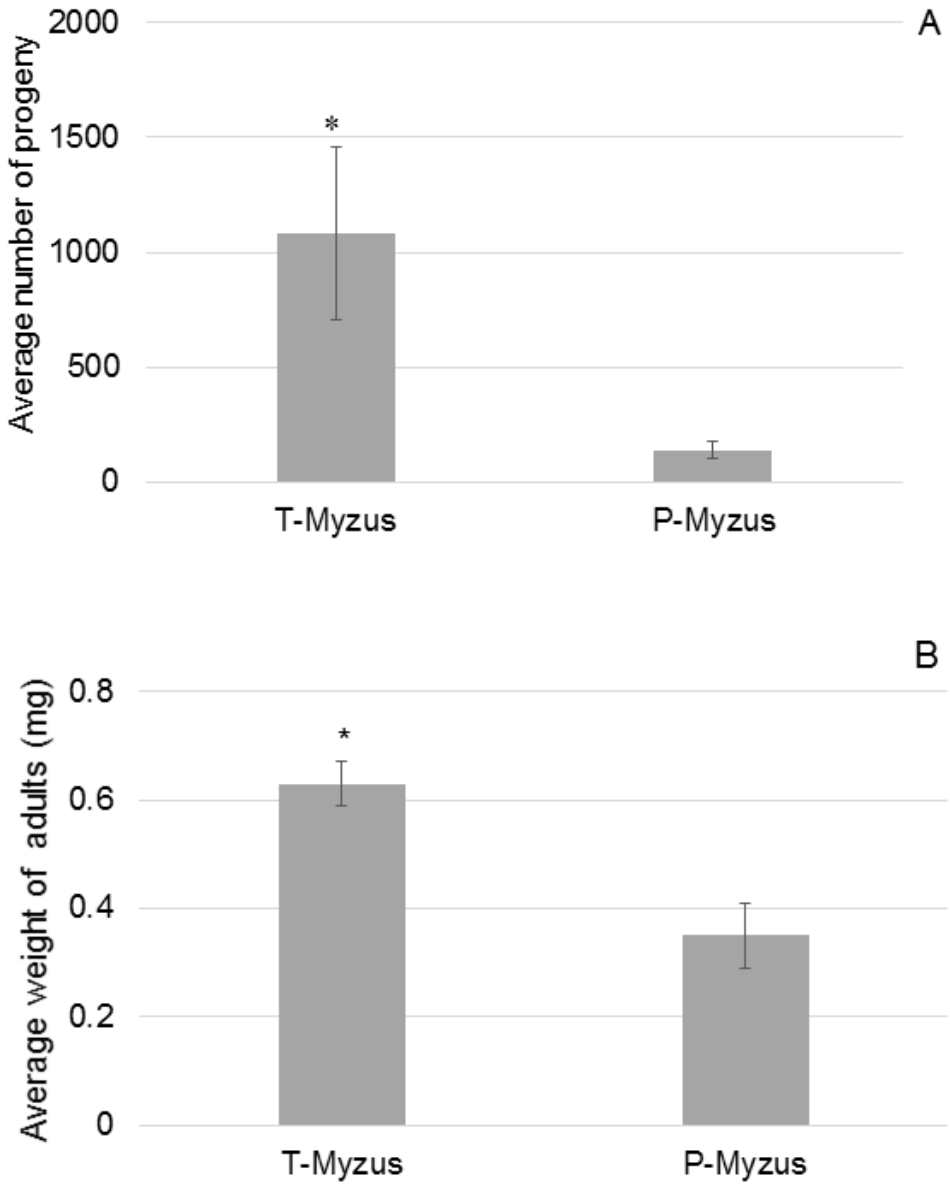

A
Average number of progeny
2000
1500
1000
500
0
*
T-Myzus
P-Myzus
B
Average weight of adults (mg)
0.8
0.6
0.4
0.2
0
*
T-Myzus
P-Myzus

Figure 4.

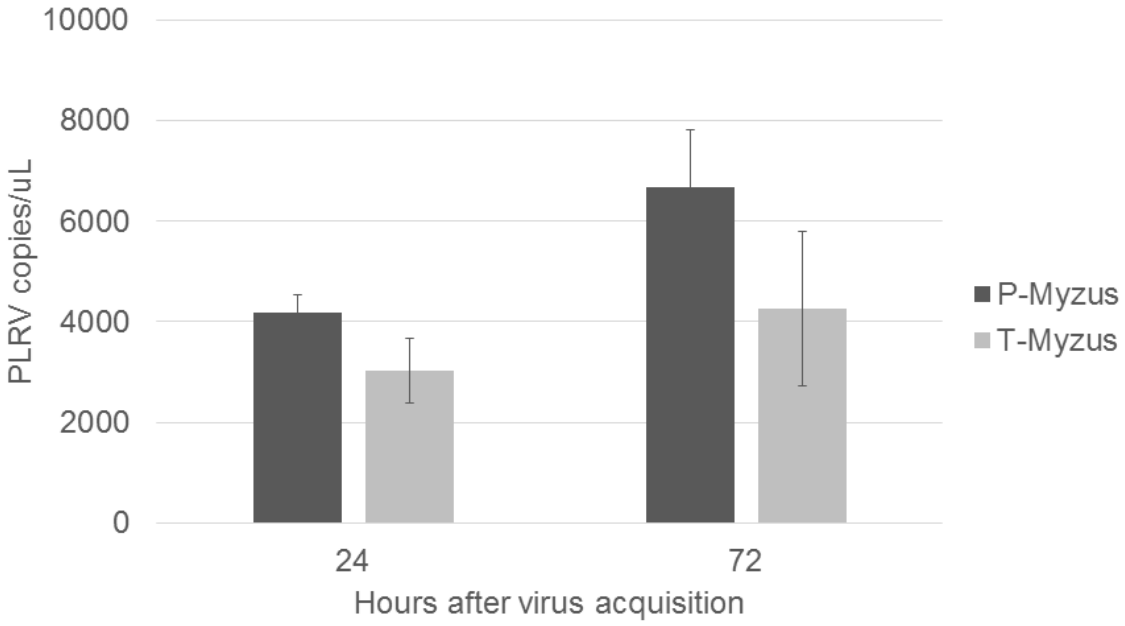

10000
8000
6000
4000
2000
0
PLRV copies/uL
24
72
Hours after virus acquisition
P-Myzus
T-Myzus

Figure 5.

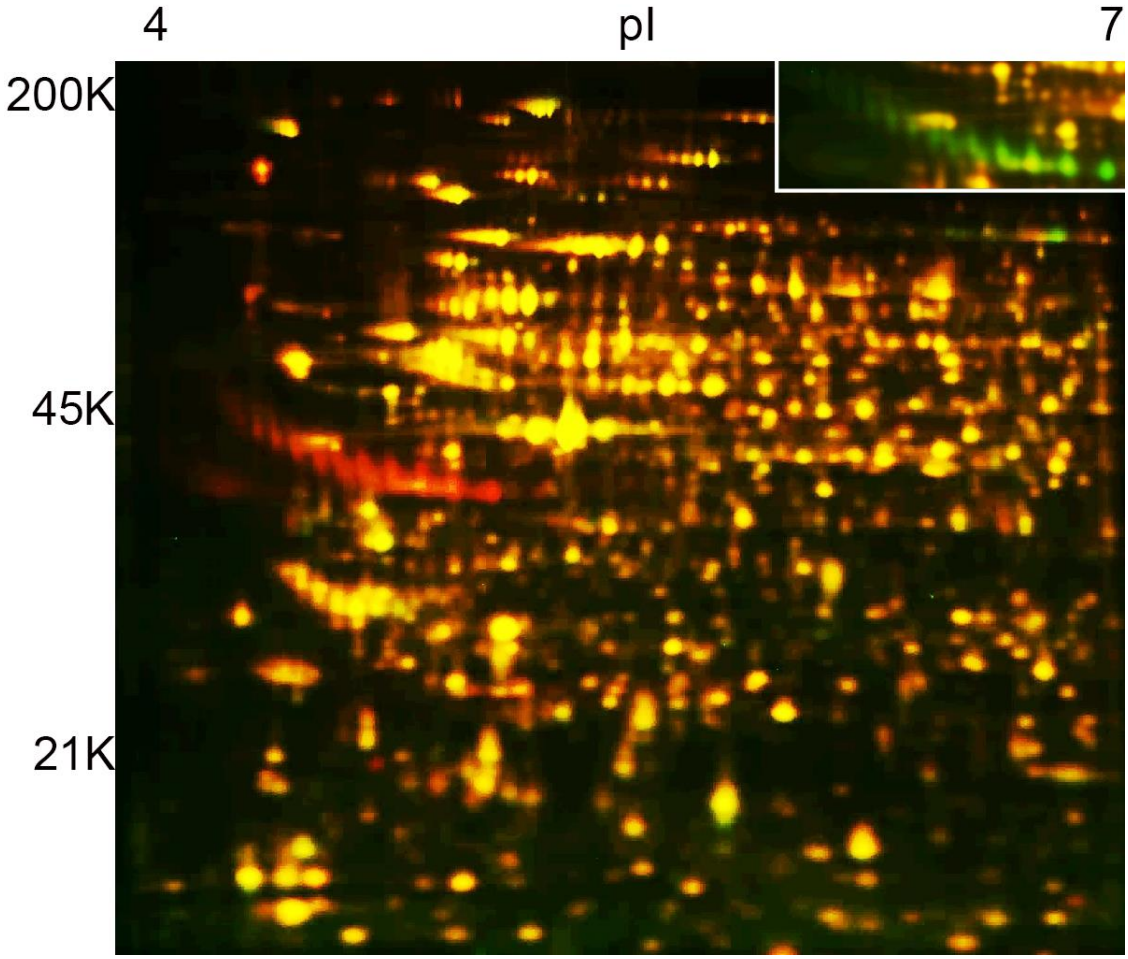


Figure 6.

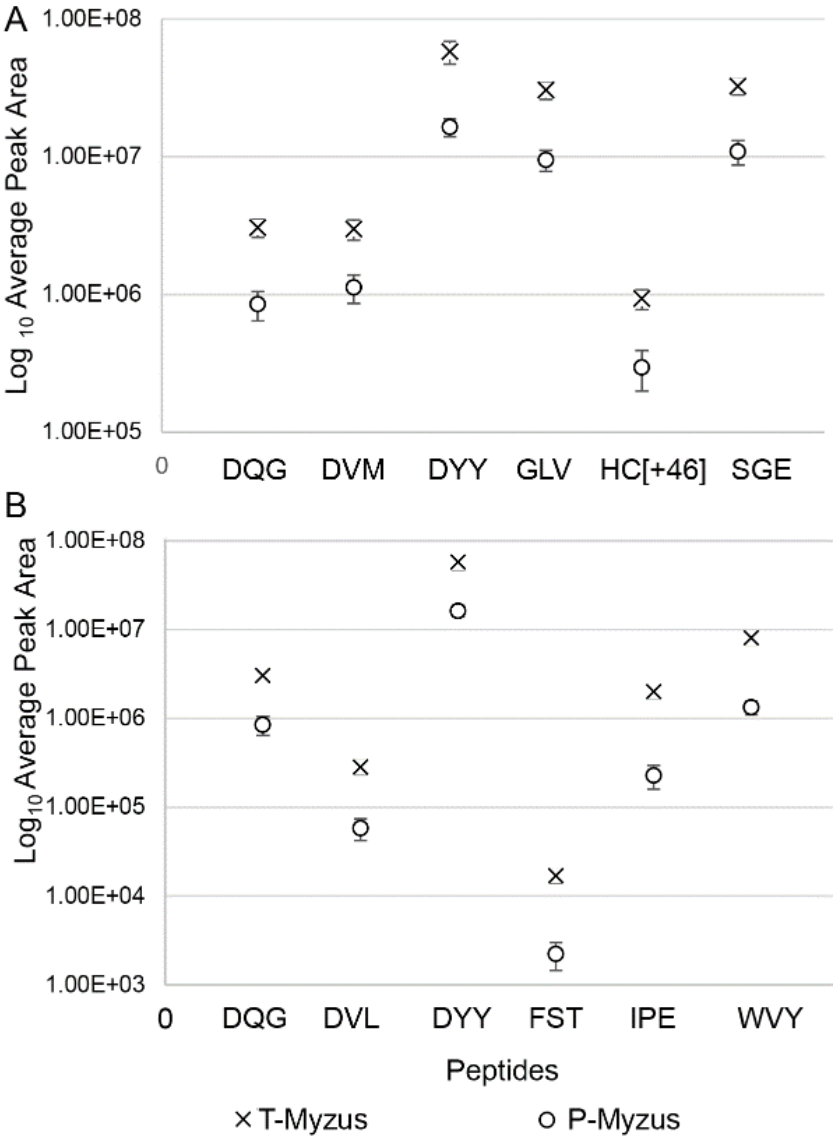

A
1.00E+08
1.00E+07
1.00E+06
1.00E+05
Log 10 Average Peak Area
0
DQG
DVM
DYY
GLV
HC[+46]
SGE
B
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
Log10 Average Peak Area
0
DQG
DVL
DYY
FST
IPE
WVY
Peptides
× T-Myzus
○ P-Myzus

Figure 7.

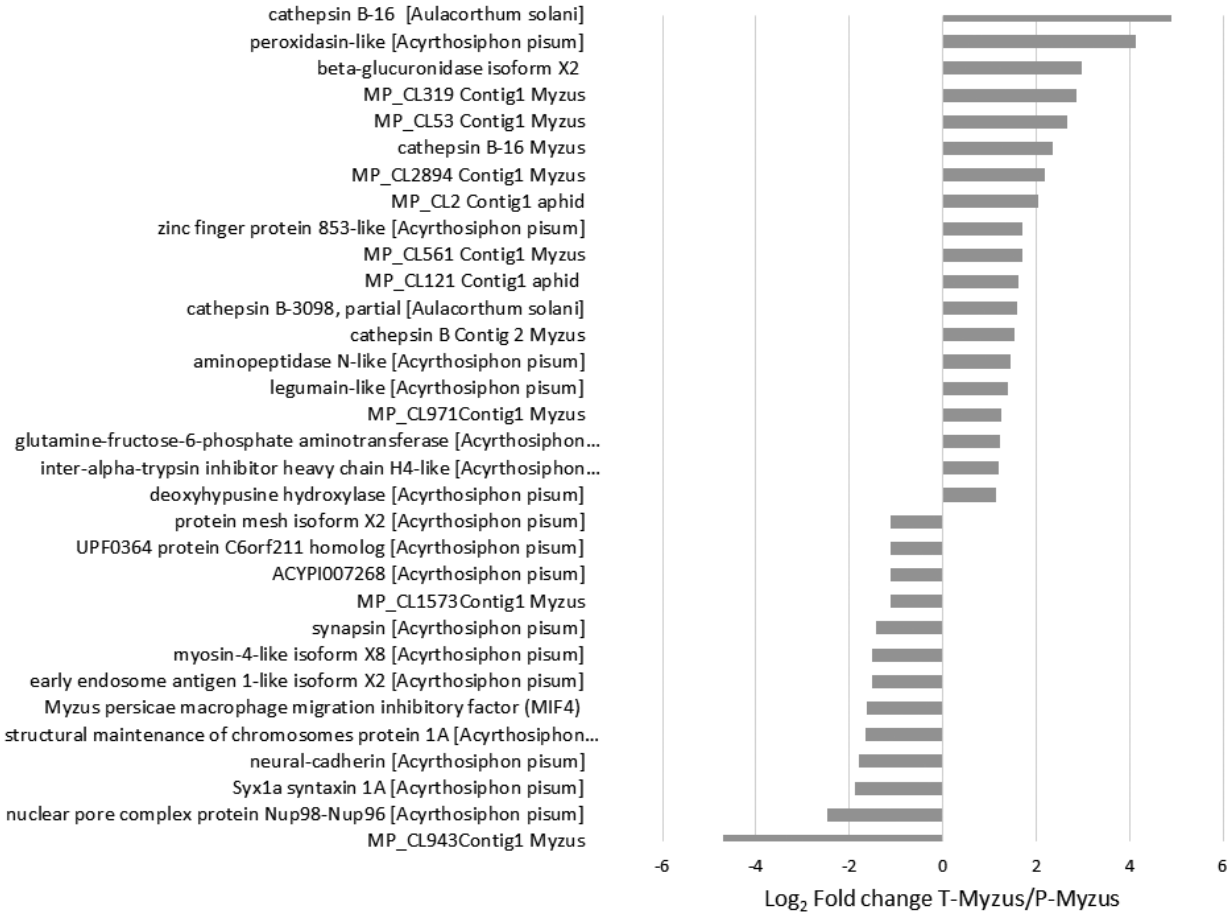

cathepsin B-16 [Aulacorthum solani]
peroxidasin-like [Acyrthosiphon pisum]
beta-glucuronidase isoform X2
MP_CL319 Contig1 Myzus
MP_CL53 Contig1 Myzus
cathepsin B-16 Myzus
MP_CL2894 Contig1 Myzus
MP_CL2 Contig1 aphid
zinc finger protein 853-like [Acyrthosiphon pisum]
MP_CL561 Contig1 Myzus
MP_CL121 Contig1 aphid
cathepsin B-3098, partial [Aulacorthum solani]
cathepsin B Contig 2 Myzus
aminopeptidase N-like [Acyrthosiphon pisum]
legumain-like [Acyrthosiphon pisum]
MP_CL971Contig1 Myzus
glutamine-fructose-6-phosphate aminotransferase [Acyrthosiphon...
inter-alpha-trypsin inhibitor heavy chain H4-like [Acyrthosiphon...
deoxyhypusine hydroxylase [Acyrthosiphon pisum]
protein mesh isoform X2 [Acyrthosiphon pisum]
UPF0364 protein C6orf211 homolog [Acyrthosiphon pisum]
ACYPI007268 [Acyrthosiphon pisum]
MP_CL1573Contig1 Myzus
synapsin [Acyrthosiphon pisum]
myosin-4-like isoform X8 [Acyrthosiphon pisum]
early endosome antigen 1-like isoform X2 [Acyrthosiphon pisum]
Myzus persicae macrophage migration inhibitory factor (MIF4)
structural maintenance of chromosomes protein 1A [Acyrthosiphon...
neural-cadherin [Acyrthosiphon pisum]
Syx1a syntaxin 1A [Acyrthosiphon pisum]
nuclear pore complex protein Nup98-Nup96 [Acyrthosiphon pisum]
MP_CL943Contig1 Myzus
-6
-4
-2
0
2
4
6
Log2 Fold change T-Myzus/P-Myzus

Figure 8.

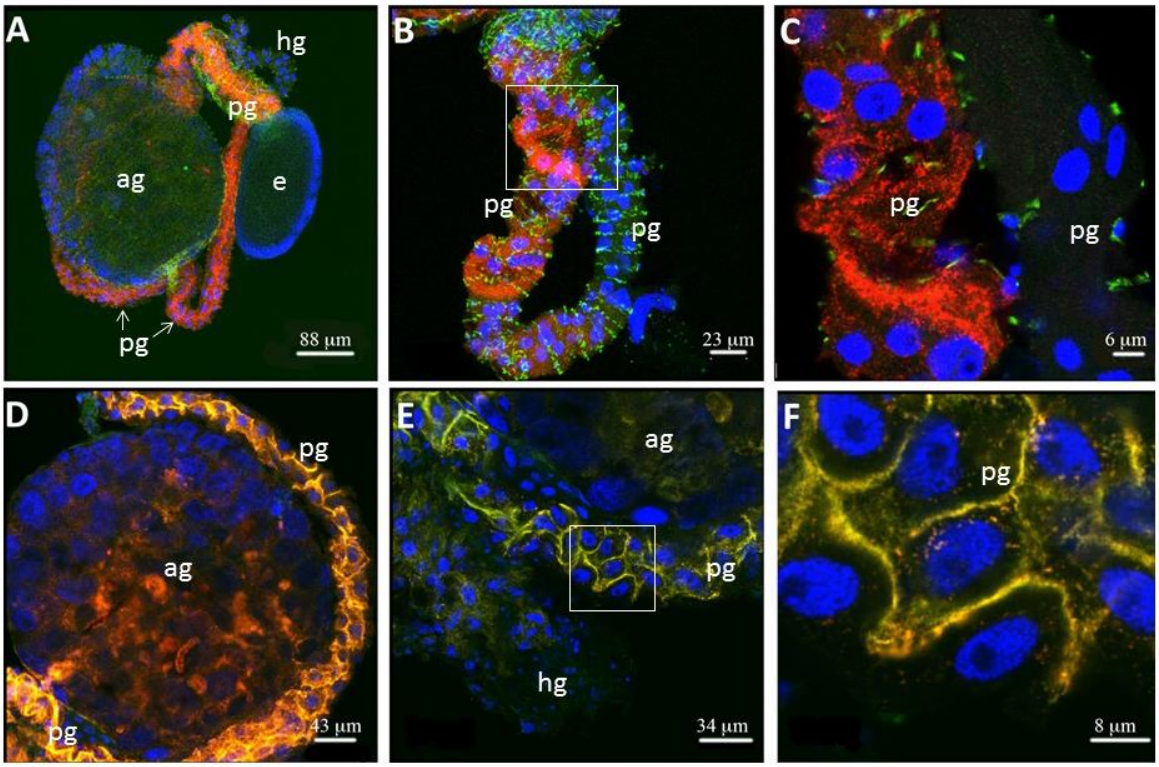

A
hg
pg
ag
e
pg
88 µm
B
pg
pg
23 µm
C
pg
pg
6 µm
D
pg
ag
pg
43 µm
E
ag
pg
hg
34 µm
F
pg
8 µm

Figure 9.

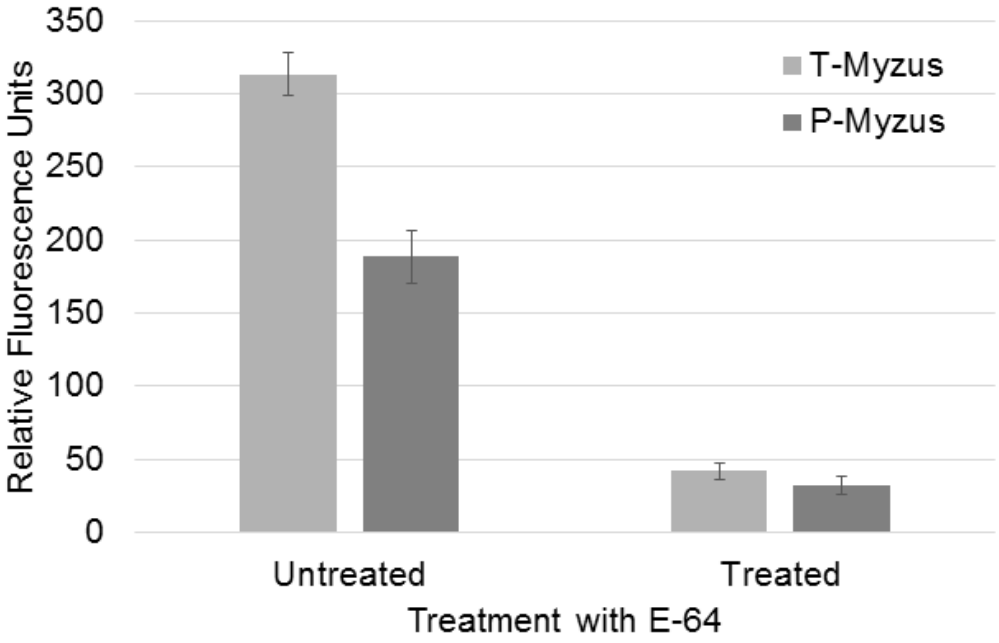

350
300
250
200
150
100
50
0
Relative Fluorescence Units
T-Myzus
P-Myzus
Untreated
Treated
Treatment with E-64

Figure 10.

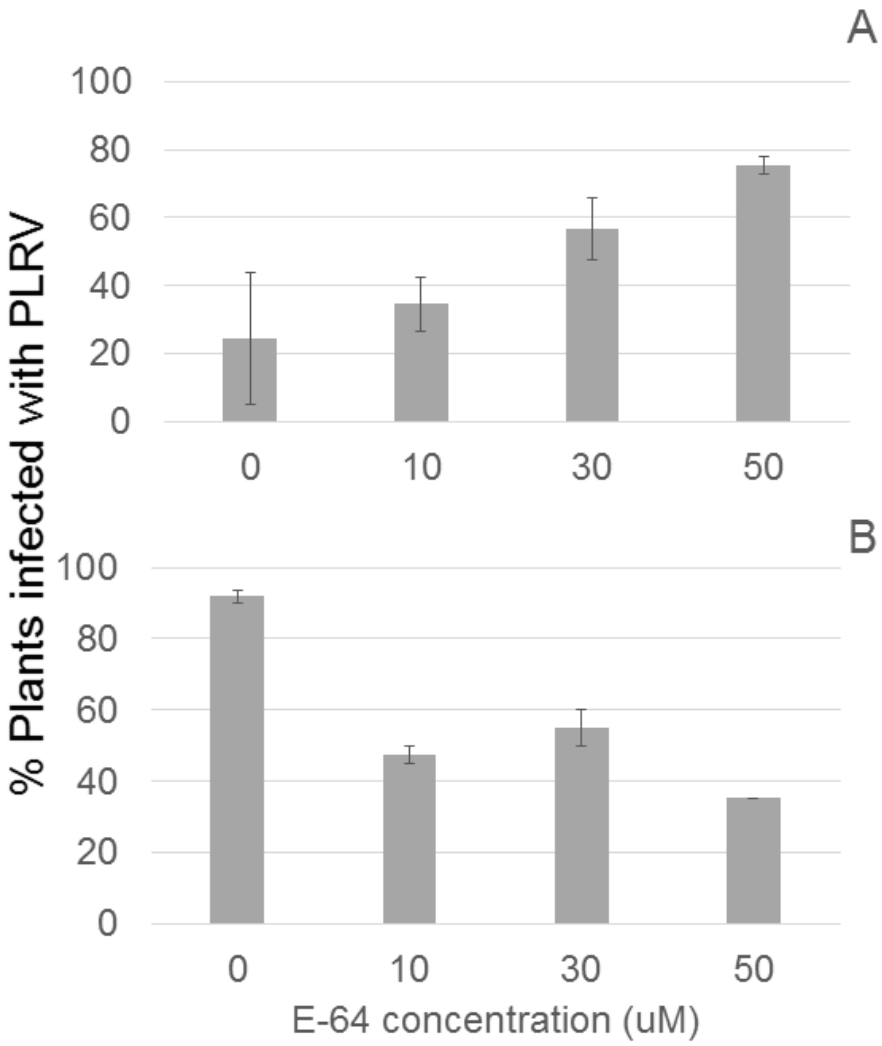

A
100
80
60
40
20
0
0
10
30
50
B
100
80
60
40
20
0
0
10
30
50
% Plants infected with PLRV
E-64 concentration (uM)

Figure 11.

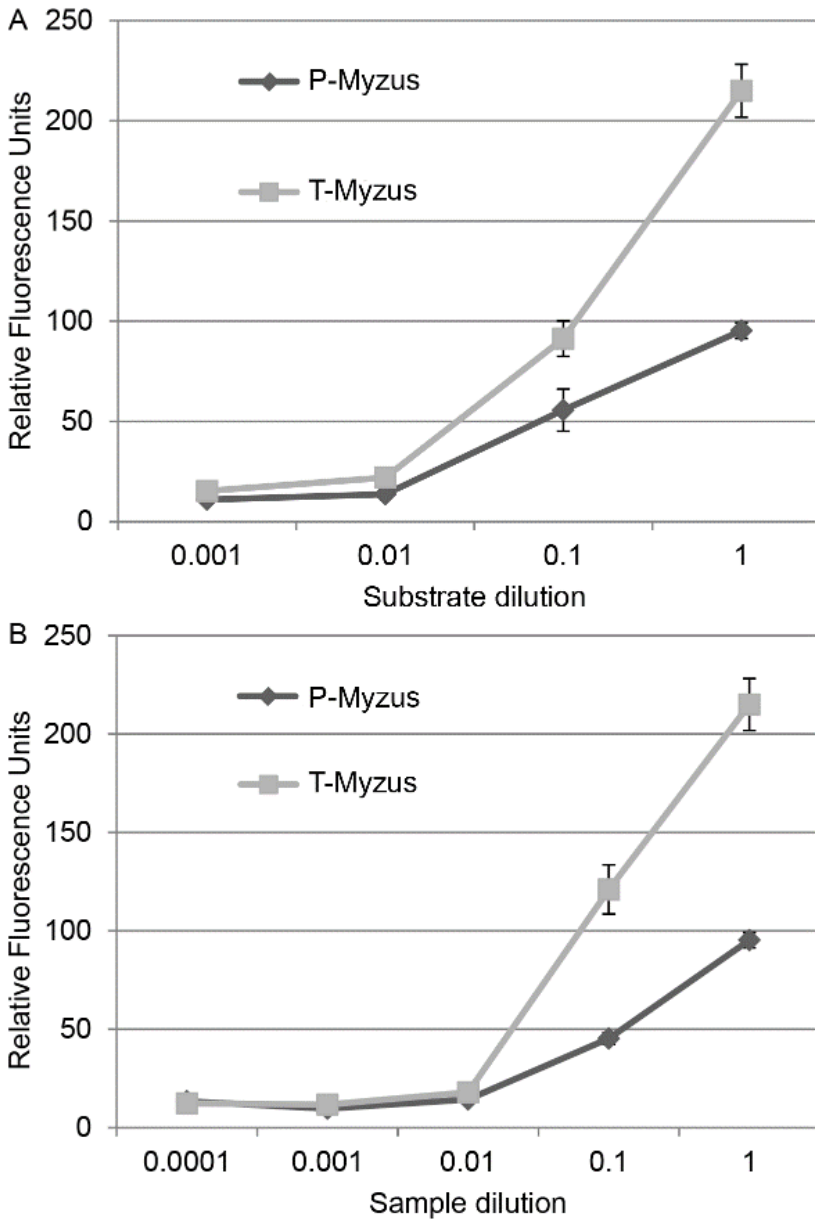

A
250
200
150
100
50
0
Relative Fluorescence Units
P-Myzus
T-Myzus
0.001
0.01
0.1
1
Substrate dilution
B
250
200
150
100
50
0
Relative Fluorescence Units
P-Myzus
T-Myzus
0.0001
0.001
0.01
0.1
1
Sample dilution

Figure 12.


1
0.8
0.6
0.4
0.2
0
Ca+ (mg/g sample)
*
T-Myzus
P-Myzus
Myzus persicae colony

Chapter 3

Is there a role for symbiotic bacteria in plant virus transmission by insects?

Published as a review article in Current Opinion of Insect Science (2015) 8: 69-78

Patricia V. Pinheiro[1,2,3], Adi Kliot[2,4], Murad Ghanim[4] and Michelle Cilia[2, 5, 6,*]

[1] Department of Entomology, Cornell University, Ithaca, NY 14853
[2] Boyce Thompson Institute for Plant Research, Ithaca, NY 14853
[3] Embrapa Rice and Beans, Santo Antônio de Goiás, 75375-000, Brazil
[4] Department of Entomology, Volcani Center, Bet Dagan 50250, Israel
[5] Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853
[6] Agricultural Research Service, Biological Integrated Pest Management Research Unit, Ithaca, NY 14853

*For correspondence: mlc68@cornell.edu

ABSTRACT

During the process of circulative plant virus transmission by insect vectors, viruses interact with different insect vector tissues prior to transmission to a new host plant. An area of intense debate in the field is whether bacterial symbionts of insect vectors are involved in the virus transmission process. We critically review the literature in this area and present a simple model that can be used to quantitatively settle the debate. The simple model determines whether the symbiont is involved in virus transmission and determines what fraction of the pathogen transmission phenotype is contributed by the symbiont. The model is general and can be applied to any vector-pathogen-symbiont interactions.

KEY WORDS

Buchnera, *Rickettsia*, vectoring efficiency, plant virus, GroEL, Hemiptera, symbiont, endosymbiont, mutualism, virus transmission, aphid, whitefly, vector, microbiota

INTRODUCTION

Vector-borne plant viruses are broadly classified by the length of time they are associated with their insect vector. These classifications are referred to as different types of virus-vector associations. Non- and semi-persistent viruses remain associated with insect vector tissues (primarily the mouthparts and

foreguts) for minutes to hours, whereas the persistent viruses stay with the insect for its entire life. Persistent viruses circulate throughout the body of the insect prior to the inoculation of a new host plant. During circulative transmission, the virus has to transverse the gut, to survive in the hemolymph and to transverse salivary tissues. Circulative viruses can be sub-classified as propagative or non-propagative, where the former replicates in the vector tissues, possibly causing pathology in the insect, and the latter moves through vector tissues with no detectable virus replication (1). Non-propagative, circulative viruses include members of the *Luteoviridae* (luteovirids)*, Geminiviridae,* and *Nanoviridae,* which are transmitted by aphids and whiteflies. In all classifications, some viruses can be transmitted by multiple insect species, implying that the association is not virus-vector species specific, while others display extraordinary vector-virus species specificity. Insect populations in many species vary in their ability to transmit plant viruses (2–9) and the effects of plant viruses and infected plants on non-vector populations within a vectoring species are not known.

Recently, the vector manipulation hypothesis has been proposed to explain the relationship between insect vectors and the plant viruses they transmit (10, 11). The hypothesis predicts that a virus will promote plant-to-plant spread by influencing the plant host selection behavior of the insect vector (11, 12) and increasing the production of alate individuals (13, 14). Aphids are more attracted to plants infected with viruses they transmit than to healthy plants or to plants infected with viruses that have other modes of transmission. This suggests that the type of virus-vector association shapes the extent to which viruses are able to

manipulate their vectors (15–21). In propagative associations, the insect vector is also a host of the virus, and virus replication can be detrimental to the insect host. Vector pathogenicity has an impact on the attractiveness of virus-infected plants to insect vectors, and consequently, on the fitness of the virus. For instance, positive or neutral effects on vector performance have been extensively reported for persistently transmitted viruses that are dependent on their insect vectors for transmission (12, 22–29). On the other hand, negative and sometimes neutral effects on insects have been reported mainly for plants infected with viruses and other pathogens that are not transmitted or not exclusively transmitted by the insect species studied (17, 30–34). Negative effects have been reported for plants infected with virus that replicate in the insect vector (35–38), which is expected as propagative viruses can have detrimental effects on their insect hosts. Altogether, these observations raise the intriguing idea that some persistent plant viruses may be mutualistic symbionts of their insect vectors. Depending on the level of specificity of the tritrophic interaction and the history of virus-vector co-evolution, viruses and insect vectors may collaborate in fighting plant immune responses (39–41), in minimizing pathogenic effects in the vector (42, 43) or in minimizing perturbations of vector biochemistry (44).

The relationship among plant viruses, insect vectors and plant hosts is hypothesized to involve a fourth party: bacterial symbionts harbored within the body of the insect vector. Primary bacterial symbionts are obligate and remain confined in specialized cells called bacteriocytes. Non-vital, secondary symbionts reside either in bacteriocytes or in other tissues such as the gut, fat body and

reproductive tissues (45). Bacterial symbioses in the Hemiptera are exemplary of symbioses in nature. For the Hemiptera, natural selection favored the evolution of symbioses to complement the insect diet with a concomitant reduction in symbiont genome size during co-evolution by the loss of symbiont genes involved in the metabolism of compounds that can be obtained from the insect host cytoplasm (46). The primary symbiont contributes mainly to the nutritional needs of the insect, providing the essential amino acids that the insect is unable to synthesize or obtain from its diet in sufficient amounts (47). The association confers an advantage to the insect, which avoids competition with other insect species by enabling hemipteran insects to feed on an exclusive niche: the phloem sap of plants, which is rich in carbohydrates and poor in other nutrients. Without the primary symbiont, the insects grow poorly and have essentially no reproductive output (48). The symbiosis is also obligate for the bacteria, which are not found in isolation in nature, and the bacteria are not culturable (49). How bacterial symbionts participate in modulating vector responses (including behavioral, biochemical, metabolic, etc.) to manipulation by plant viruses is largely unknown. Whereas there is a wealth of literature supporting the hypothesis that the insect vector's microbiota is involved in the transmission of insect-borne, animal-infecting viruses (50), there are contrasting reports from the literature regarding whether symbiotic bacteria play a role in the transmission of plant viruses by insect vectors.

Aphids and the primary symbiont, *Buchnera*

One of the best-studied, classical examples in the plant vector biology field is the symbiosis between aphids and the γ-proteobacterium *Buchnera aphidicola.* *Buchnera* cells reside in bacteriocytes. A detailed proteomics analysis of the pea aphid (*Acyrthosiphon pisum*) bacteriocyte showed the expression of metabolite transport proteins and all enzymes involved in amino acid metabolism (51). In the pea aphid bacteriocyte, aphid-derived, cysteine-rich, secreted proteins are expressed (52). Intriguingly, these proteins resemble cysteine-rich, secreted proteins that are expressed in the specialized cells of leguminous plants that harbor nitrogen-fixing bacteria, suggesting convergence of symbiotic molecular signatures across highly diverse taxa (47, 52). Sequencing of the *A. pisum* genome revealed that aphids have lost the genes that other insect taxa would normally leverage to mount an immune response against bacteria (53, 54) and an expansion of the genes involved in the production of short, non-coding RNAs (55). Highly specific interactions between invertebrate host immune systems that foster relationships with invertebrate bacterial symbionts are wide-spread in nature and vary with each invertebrate host-symbiont pair and function of the symbiosis (56).

Since the early 1990s, a central hypothesis about the regulatory mechanism of circulative virus transmission by aphids proposed that *Buchnera* symbionin, the most abundant *Buchnera* protein (51, 57) and one of the most abundant proteins in the aphid proteome (57), protected luteovirids from degradation during transit in the aphid hemolymph, or insect blood (58). *Buchnera* symbionin is a homolog of the *Escherichia coli* GroEL protein sharing greater than 80% sequence identity with its *E. coli* homolog. Symbionin will be referred to hereafter as GroEL. To

understand this hypothesis, a more detailed description of the proposed circulative pathway in the aphid (1, 42, 59) is required. First, a feeding aphid ingests virus particles during phloem ingestion together with phloem sap proteins, RNAs and other metabolites. Once in the insect, the virus must be acquired. Virus acquisition involves the passage of virus through either the midgut or the hindgut via an endocytosis mechanism (1). From the gut, the virus passes into the hemocoel, or body cavity. Different species of circulative viruses have different tropisms for sections of the gut that are determined solely by the topology of the virus capsid (60–67). By a poorly understood mechanism, viruses are proposed to move in the hemolymph to the salivary tissues where they enter the cells through endocytosis and are inoculated into a new host plant during feeding. It is during hemolymph transport where GroEL has been proposed to protect the virus while passing through the insect hemolymph. The supposed, GroEL-protected luteovirids are then transported across the paired, accessory salivary glands. Whitefly-transmitted begomoviruses and the aphid-transmitted nanovirus *Banana bunchy top virus* (BBTV) are transported across the primary salivary glands (1, 68).

Evidence supporting the involvement of GroEL in the circulative transmission of the luteovirids *Potato leafroll virus* (PLRV) and species of *Barley/Cereal yellow dwarf viruses* (B/CYDV) was gleaned from several studies. According to the authors, treatment with antibiotics to kill *Buchnera* in the aphid reduced luteovirid transmission by aphids by 70% (58). *Buchnera* GroEL binds to PLRV (48) and two species of BYDV *in vitro* (69, 70). GroEL protein was also found in the aphid hemolymph (58) and saliva using immunostaining with a polyclonal,

anti-GroEL antibody, although it was not shown if the protein was associated with virus (69) in the saliva. Recently, *Buchnera* GroEL has also been detected in aphid saliva using proteomics approaches (71, 72). Care should be taken to pinpoint the origin of GroEL found within aphid saliva using antibodies and mass spectrometry because of the difficulties in unambiguously distinguishing between *Buchnera* GroEL and GroEL expressed from other bacteria present on the aphid stylet and introduced into the diet.

The well-characterized biochemical properties of GroEL are relevant to its putative role in virus transmission. GroEL binds to many proteins with very diverse linear and structural motifs (73), a property of the protein that is critical for its function as a protein-folding chaperone. Additionally, this protein is notoriously immunogenic, such that very small amounts of GroEL from bacteria of many different taxa would cross-react with polyclonal antiserum raised against GroEL proteins from other bacterial species (74, 75).

Recently, the involvement of GroEL in luteovirid transmission has been called into question. The proposed GroEL binding site on luteovirids is the readthrough domain (RTD), a viral structural protein incorporated into virus particles. Regions of the RTD are exposed on the surface of virions (67) and found to interact with GroEL (76). The RTD-GroEL interaction has been proposed to provide stability to luteovirids in the aphid hemolymph. Contrary to the hemolymph stability model, Liu and colleagues (2009) showed the stability of a naturally occurring mutant of *Pea enation mosaic virus* that lacked the RTD was the same as wild-type virus in the aphid hemolymph (77). Another layer of evidence against

the involvement of GroEL in virus transmission came from the work of Bouvaine and colleagues in 2011. They showed that GroEL protein was restricted to the bacteriocyte using an anti-GroEL monoclonal antibody specific to *Buchnera* GroEL protein, a reagent that was critical to the success of their overall approach. GroEL was never detected in the hemolymph, gut or fat body, which makes it highly unlikely for a luteovirid-GroEL protein interaction to occur *in vivo* (70). The specific detection of GroEL in hemolymph in previous reports was likely dependent upon the method used for aphid dissection. Van den Heuval and colleagues used cornicle amputation to extract hemolymph from the green peach aphid, the vector for PLRV (58). However, Bouvaine and colleagues showed that in the pea aphid GroEL was not detected in hemolymph if leg amputation is used for hemolymph collection but it could be detected in hemolymph using cornicle amputation (70). In an elegant, *in vitro* binding study coupled to structural modeling, the virus-binding site in GroEL was mapped and coincided with the GroEL multimer interfacial region (70). The low stoichiometry of virus:symbionin in a viruliferous aphid would make virus access to and competition for the binding site unlikely. Other studies have used genetics coupled to proteomics analysis for investigating aphid and *Buchnera* proteins involved in luteovirid transmission. The expression of the GroEL protein was not correlated with the CYDV-vectoring phenotype in an F2 population and field-collected populations of the greenbug (*Schizaphis graminum*) (78–80).

Proteomics and genetics experiments conducted on the greenbug suggest an alternative to the GroEL model for the involvement of *Buchnera* in virus transmission. Cilia and colleagues described an association between a particular

genotype of *Buchnera* and the ability to efficiently transmit the RPV strain of CYDV. Cilia and colleagues observed a tight correlation between aphid genotypes in F2 and field-collected populations and a *Buchnera* genotype, where one *Buchnera* genotype is required for virus transmission (80). In the F2 and field-collected aphid populations, proteins involved in the transmission of the RPV species of CYDV were quantified and characterized using mass spectrometry. Efficient vector genotypes expressed proteins derived from vectoring alleles that segregated with efficient vectoring ability of CYDV-RPV. Two distinct genotypes of *Buchnera* were observed to segregate in the F2 population, indicating genetic heterogeneity of the aphid female used in the cross. With no exception, all greenbug clones and field-collected biotypes that efficiently transmitted CYDV-RPV harbored the same genotype of *Buchnera* (7, 79, 80) but the direction of causality remains an unsolved puzzle. It is pertinent to note that in the greenbug, efficient vectors are more commonly found associated with wild grass species whereas non-vectors that do not transmit virus efficiently are found more commonly associated with cultivated crops (81), raising the idea that in the greenbug, variation in symbiont genotype may be a function of plant host adaptation.

Two studies also raise the intriguing idea that the aphid microbiota, including *Buchnera*, influences the aphid-plant relationship during insect feeding and plant colonization (71, 82). Such an influence may foreseeably impact virus transmission and plant infection. This alternative hypothesis predicts the aphid microbiota has an impact on plant virus transmission not via protein interactions within the vector, but via indirect effects of modulation of plant defense networks

that influence virus transmission or plant infection. One take on this idea was presented in a paper by Chaudhary and colleagues (71) that is, *Buchnera* GroEL from the green peach aphid (*Myzus persicae*) “betrays the aphid” and triggers plant defenses which have a negative impact on aphid fitness. Elzinga and colleagues made a parallel discovery when they used transgenic *Arabidopsis thaliana* plants expressing *Buchnera* GroEL under the control of a phloem-specific promoter (82). *Buchnera* GroEL induced plant defense responses commonly mounted against bacterial infection, which subsequently had a negative impact on green peach aphid fitness. Importantly, in addition to the origin of the GroEL detected in saliva not being definitively confirmed (as discussed above), experiments to test why *Buchnera* would have been favored by natural selection to betray its obligate aphid host, as suggested by Chaudhary and colleagues, and trigger plant defense responses were not conducted in either study. In the event that GroEL is present in aphid saliva, care should be taken in the interpretation of the phenotypic effects reported in both studies because the levels of GroEL in plants modified to express the protein *in planta* would be much higher than in plants exposed to the tiny amounts of saliva secreted during the E1 phase of feeding. That said, the local concentrations of aphid proteins in saliva *in planta* are not known and may be high at the local site of aphid feeding. So while the results obtained with *in planta* expression may not be relevant to the effects of aphid-delivered GroEL on the scale of an entire plant, local induction of plant defense pathways triggered by any aphid salivary effector protein that subsequently leads to a systemic plant response is, of course, quite likely.

Taken together, these studies do not prove or disprove that *Buchnera* is involved in luteovirid transmission but they do weaken the model where GroEL directly involved in virus movement in the aphid hemolymph and instead point to an insect genotype by symbiont genotype interaction or a completely different model involving *Buchnera* modulating plant defense against aphids. To wit, Watanabe and Bressan showed evidence of a hemolymph-independent virus translocation path from the anterior midgut to the salivary glands in the banana aphid (*Pentalonia nigronervosa*) for BBTV (68), so a model where GroEL protects the virus during hemolymph translocation may not be accurate for some circulative viruses. An overall weakness of the *Buchnera* GroEL work to date is that different virus-vector species combinations have been used in all these studies. Although there are unifying principles in the circulative transmission pathways of aphid-transmitted plant viruses (1), it is not likely that all the molecular-level interactions regulating virus transmission in one aphid species for one virus species can be generalized for all of the luteovirids and their aphid vectors.

Whiteflies and secondary symbionts

Whiteflies, including *Bemisia tabaci,* harbor a primary symbiont, the γ-proteobacterium *Portiera aleyrodidarum*, in specialized bacteriocytes. Several genome sequencing projects of *Portiera* revealed genes involved in essential amino acid synthesis not present in the insect genome (83, 84). In addition to the primary symbiont, several secondary symbionts have been reported in whitefly

populations surveyed worldwide, including *Arsenophonus, Hamiltonella, Wolbachia, Rickettsia, Cardinium, Fritschea* and *Orientia* (85–87). These secondary symbionts are facultative residents and play more diverse roles in the association with whiteflies, for example by conferring resistance to insecticides (88, 89) and high temperatures (90) and by protecting the insects from natural enemies (91–94). Negative effects of secondary symbionts have also been reported. These negative effects may be due to a shorter evolutionary history of interaction compared to the primary symbiont or because the symbiont-insect interaction was not assessed under conditions where the symbiont would be beneficial (88, 95). Pertinent to this review, secondary symbionts are hypothesized to facilitate circulative plant virus acquisition, retention and transmission by whiteflies (5, 6, 8).

B. tabaci is a complex of 11 well-defined groups comprised of more than 30 morphologically indistinguishable species (96). Some of these species, in particular the B and Q biotypes, are highly invasive and well studied. No correlation was found between a certain *B. tabaci* species and secondary symbionts on a global scale because the same species may harbor a certain bacterium in one geographical area and not in the other (97–99). Such is the case with *Hamiltonella*; in Israel, populations of the B species carry *Hamiltonella* while Q populations do not. The populations of the B species in Israel are efficient vectors of begomoviruses mainly *Tomato yellow leaf curl virus* (TYLCV), whilst populations of the Q species in Israel tend to be poor vectors (6). These observations are in stark contrast to the situation in China (100). Studies on association between *B. tabaci* populations and the *Hamiltonella* symbiont in Israel

and China are in line with the previously described greenbug studies (7, 79, 80) that showed correlations between certain symbiont genotypes and the insect's ability to efficiently transmit virus.

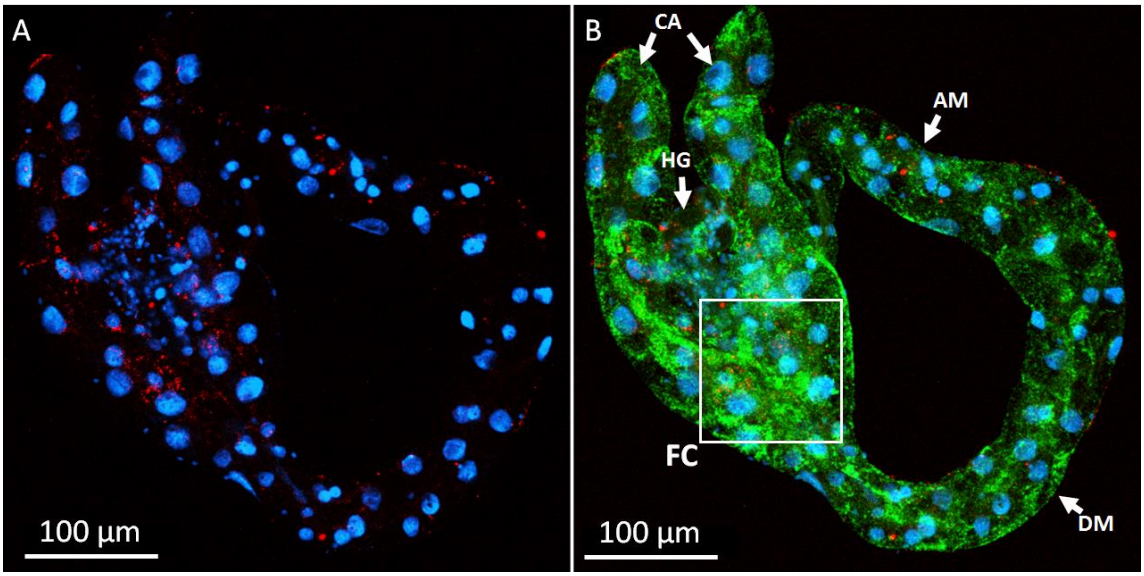
Following van den Heuvel's work with *Buchnera* GroEL in the green peach aphid, experiments were performed to test the hypothesis that symbiont-derived GroEL played a role in the transmission of TYLCV by *B. tabaci*. Similar to the findings of Bouvaine and colleagues, GroEL in *B. tabaci* localized to the bacteriocytes (101), although interestingly, TYLCV transmission by *B. tabaci* that were fed on an anti-*Buchnera* GroEL antiserum prior to virus acquisition was reduced by more than 80%. GroEL antiserum was acquired into the hemolymph, reduced the stability of the virus *in vivo* and had no effect on the stability of TYLCV in the diet used to deliver the virus (88). Later work suggested that the GroEL from *Hamiltonella* interacts specifically with TYLCV, and an immunocapture PCR with an anti-GroEL antibody and primers specific to the coat protein of TYLCV showed that the GroEL protein will co-immunoprecipitate with the virus from viruliferous insects (82). Although extensive data have been presented to support a role for GroEL in TYLVC transmission, in light of the work of Bouvaine and colleagues (70) and localization of *Hamiltonella* GroEL to bacteriocytes (101), more studies are required to prove an unambiguous, physical, protein-protein interaction between the TYLCV coat protein and *Hamiltonella* GroEL *in vivo*. Collectively though, the evidence for the involvement of *Hamiltonella* in TYLCV transmission by whiteflies is compelling. Even setting aside any putative involvement of GroEL, whiteflies of the same species without *Hamiltonella* transmit TYLCV less efficiently (8, 102)

suggesting an interaction between the presence of *Hamiltonella* in *B. tabaci* for TYLCV transmission. Su and colleagues showed that whiteflies harboring *Hamiltonella* transmit TYLCV more efficiently and accumulate more TYLCV when exposed to infected plants (8) by comparing two Q biotype lines from the same genetic background where one line retained *Hamiltonella* (H^+) and the other did not (H^-). As *in vivo* approaches for protein interaction identification and functional genomics mature for the Hemiptera, a role for GroEL in virus transmission by whiteflies and aphids may be addressed more directly and conclusively.

An additional example in which a secondary symbiont was shown to influence virus transmission is *Rickettsia* from *B. tabaci*. Unlike other symbionts, *Rickettsia* cells are located outside bacteriocytes and infect all organs of the insect, replicating to high levels in the midgut (Figure 3.1), hemolymph and salivary glands, which are known to be important sites where circulative viruses pass while circulating in the insect. Recent studies have shown that *Rickettsia* can induce both negative and positive effects on different life history parameters of *B. tabaci* (88, 103–106). Using a similar strategy as Su and colleagues (8), Kliot and colleagues produced two iso-female, genetically identical lines of the B species, one *Rickettsia*-infected ($Rick^+$) and the other uninfected ($Rick^-$) over 300 generations and showed that the presence of *Rickettsia* is associated with improved TYLCV acquisition, transmission and retention in the insect (5). Both microbes are localized in the *B. tabaci* midgut (Figure 3.1 and see (5)). Experiments showed that upon acquisition of the virus, *Rickettsia* levels dropped in the insects and only increased again after the virus titer diminished (5). In

addition, previous studies found that *Rickettsia* GroEL and TYLCV do not bind to each other (6), further supporting a GroEL-independent involvement for this symbiont. The hypothesis is that the presence of *Rickettsia* somehow primes the insect vector for virus acquisition. Alternatively, *Rickettsia* may act alone or in concert with other symbionts to influence the plant-insect interaction with concomitant effects on virus acquisition and transmission. Both the Su (8) and Kliot (5) studies demonstrated the power of using whiteflies to disentangle the role of symbionts in virus transmission by controlling for variation in the insect genetic background.

Figure 3.1. *Rickettsia* and *Tomato yellow leaf curl virus* (TYLCV) in *Bemisia tabaci* visualized using fluorescence *in situ* hybridization shows both are present in the insect midgut but do not co-localize. **A**. DAPI-stained midgut with TYLCV visualized using a cy5-labeled probe (red) specific to the TYLCV coat protein. **B.** The same midgut as in A but now visualizing *Rickettisa* cells with a cy3-labeled probe (green) specific to *Rickettsia*. TYLCV is concentrated in the filter chamber, the site where the virus moves into the insect hemolymph. The different tissues indicated on panel B: AM, ascending midgut; CA, caeca: DM, descending midgut; FC, filter chamber (area indicated by white frame); HG, hindgut. Rickettsia and TYLCV.



A quantitative model to calculate insect symbiont and insect vector contributions to pathogen transmission by insects

We propose a simple and unbiased model that can be used empirically to test whether symbionts are involved in circulative plant virus transmission and to determine their relative importance in the transmission phenotype. The simple model can be applied to any insect vector-pathogen pair and can be applied to settle the debate quantitatively. The model cannot provide information as to who benefits, whether the effects of the symbiont are direct or indirect or the molecular underpinnings of the symbiont contribution. The model only provides insight into the contribution of the symbiont relative to the vector towards transmission.

The components of the model include the insect vector, a helper symbiont genotype that is defined as one where the virus transmission efficiency is higher when it is present, and the probability of transmission. Helper symbiont genotypes can include different genotypes of primary symbionts or different secondary symbiont species and/or genotypes.

1. Let p = the probability of transmission by a vector without the helper symbiont genotype.
2. Let p' = the probability of transmission by a vector with the helper symbiont genotype.
3. The contribution of the symbiont, h, is equal to p'-p.

4. To quantify the relative importance of the symbiont to transmission efficiency, let R, the relative importance of the symbiont to transmission equal to h/(p+h).

Next we will apply the model to the whitefly- *Rickettsia and Hamiltonella* interactions reported by Kliot, Su and colleagues using the data on their $Rick^+$ and $Rick^-$ and H^+ and H^- strains, respectively. In the Kliot et al. study, the probability of TYLCV transmission for $Rick^+$ insects (p') was 0.79 and the probability of transmission for $Rick^-$ insects (p) was 0.41. Thus, the contribution of *Rickettsia* to transmission is equal to 0.38. The fraction of the transmission efficiency for these isofemale lines provided by *Rickettsia* is equal to 0.48. These data indicate that *Rickettsia* contribute substantially to the TYLCV transmission efficiency of that particular isofemale line, but not as much as the insect vector contribution. The Su et al. data allow us to measure the impact of *Hamiltonella* on TYLCV transmission as a function of insect densities. At densities of one, five or ten viruliferous females placed on a healthy tomato plant, the contributions of *Hamiltonella* are 0.6, 0.9, and 0.7, respectively, and the fractions of the TYLCV transmission are 1.0, 0.9, and 0.7, respectively. What these data show is that at very low densities, the importance of the symbiont to the ability of the insect to transmit virus is the highest. With increasing numbers of conspecifics, each insect transmitting virus at a lower level can compensate for not having the symbiont, somewhat, because of an increasing chance that at least one individual will transmit the virus. Tying this story back to the vector manipulation hypothesis, the Q biotype, which harbors *Hamiltonella* in China performs better on TYLCV-infected plants than the B

biotype, which does not harbor *Hamiltonella* in China (100). In the H^+ strain, the performance benefit was shown to be the result of suppression of the anti-herbivore jasmonic acid plant defense pathway (107), data that provide unambiguous support for the hypothesis that symbiotic bacteria influence the plant-insect relationship. Excitingly, suppression of the jasmonic acid pathway was shown to be dependent on cross-talk with the anti-microbial salicylic acid signaling pathway (107). Taken together, these data in the context of this simple model may explain how the Q biotype rapidly spread in China. Whiteflies able to obtain *Hamiltonella* would transmit virus more efficiently as individuals and be afforded a fitness benefit on virus-infected plants. Aided by the ability to suppress plant defenses (which could impart a benefit to the virus as well), *Hamiltonella* would enable infected individuals to propagate more efficiently than their conspecifics. A rare whitefly genotype with such an advantage would be predicted to spread very rapidly throughout a sexually reproducing population.

Summary and Conclusions

Critical review of the literature revealed that further study is needed before concepts regarding the role of bacterial symbionts in plant virus transmission can be generalized across different vector-virus pathosystems. Differences may exist between primary and secondary symbionts regarding their involvement in circulative virus transmission, and these differences may be context, host, and genotype (any of the parties) dependent. Future work should focus on whether

symbionts influence the vector-plant relationship and if so, whether these impacts are direct or indirect and represent cooperation or conflict among the parties involved. Proof of a direct involvement would require a symbiont protein involved in protein complex formation with a virus protein. Indirect involvement would be demonstrated when a symbiont influences some aspect of the insect vector physiology that changes its ability to transmit virus. Molecular studies should be aimed at understanding how symbionts regulate circulative virus transmission and should also be focused on specific barriers for virus movement (gut, salivary glands, and hemolymph). Caution should be taken when interpreting experiments that involve bacterial proteins in insect vectors, as there is a high level of protein homology across many bacterial species that could reside within an insect. Tools must be carefully developed and tested for molecular analysis. A deeper understanding of how plant viruses and symbionts promote changes in vector behavior is needed, as are tests for the involvement of the insect vector microbiota in this process. Advances in new technologies for studying organismal biology, ‘omics, functional genomics and protein interactions will undoubtedly facilitate research progress in this area for plant virus vectors.

Acknowledgements

This work was supported by an Association of Independent Plant Research Institutes Research Award, National Science Foundation grants 1354309 and

1109989, BARD Graduate Student Fellowship # GS-27-2014, and an Embrapa Graduate Student Fellowship. The funding sources had no role in the analysis and interpretation of data, in the writing of the report, or in the decision to submit this article for publication.

References

1. **Gray S**, **Cilia M**, **Ghanim M**. 2014. Circulative, “nonpropagative” virus transmission: an orchestra of virus-, insect-, and plant-derived instruments. Adv Virus Res **89**:141–199.

   This extensive review on circulative, non-propagative virus transmission discusses different aspects of virus transmission and all new discoveries in the field taking into account not only the virus and plant side of the story, but also the vector mechanisms involved in facilitating virus transmission.

2. **Burrows ME**, **Caillaud MC**, **Smith DM**, **Gray SM**. 2007. Biometrical genetic analysis of luteovirus transmission in the aphid Schizaphis graminum. Heredity **98**:106–113.

3. **Papura D**, **Jacquot E**, **Dedryver CA**, **Luche S**, **Riault G**, **Bossis M**, **Rabilloud T**. 2002. Two-dimensional electrophoresis of proteins discriminates aphid clones of Sitobion avenae differing in BYDV-PAV transmission. Arch Virol **147**:1881–1898.

4. **Bourdin D**, **Rouze J**, **Tanguy S**, **Robert Y**. 1998. Variation among clones of Myzus persicae and Myzus nicotianae in the transmission of a poorly and a highly aphid-transmissible isolate of potato leafroll luteovirus (PLRV). Plant Pathol **47**:794–800.

5. **Kliot A**, **Cilia M**, **Czosnek H**, **Ghanim M**. 2014. Implication of the bacterial endosymbiont Rickettsia spp. in interactions of the whitefly Bemisia tabaci with tomato yellow leaf curl virus. J Virol **88**:5652–5660.

   In this article, the authors present evidence for the involvement of the Rickettsia secondary symbiont of Bemisia tabaci in transmission of TYLCV.

A Rickettsia-infected strain of B. tabaci displayed higher acquisition, transmission and retention of TYLCV than a genetically identical strain lacking this symbiont. The study also provided evidence for accumulation of Rickettsia and TYLCV in the whitefly midgut.

6. **Gottlieb Y**, **Zchori-Fein E**, **Mozes-Daube N**, **Kontsedalov S**, **Skaljac M**, **Brumin M**, **Sobol I**, **Czosnek H**, **Vavre F**, **Fleury F**, **Ghanim M**. 2010. The Transmission Efficiency of Tomato Yellow Leaf Curl Virus by the Whitefly Bemisia tabaci Is Correlated with the Presence of a Specific Symbiotic Bacterium Species. J Virol **84**:9310–9317.

   The authors of this study hypothesize that the GroEL protein produced by the secondary endosymbiont Hamiltonella, present in the B biotype of B. tabaci, but not in the Q biotype, facilitates the transmission of TYLCV. The authors show that the Hamiltonella protein in the B biotype binds to TYLCV in vitro and that B. tabaci populations that harbor Hamiltonella are more efficient vectors than those that do not harbor the symbiont.

7. **Pinheiro P**, **Bereman MS**, **Burd J**, **Pals M**, **Armstrong S**, **Howe KJ**, **Thannhauser TW**, **MacCoss MJ**, **Gray SM**, **Cilia M**. 2014. Evidence of the Biochemical Basis of Host Virulence in the Greenbug Aphid, Schizaphis graminum (Homoptera: Aphididae). J Proteome Res **13**:2094–2108.

8. **Su Q**, **Pan H**, **Liu B**, **Chu D**, **Xie W**, **Wu Q**, **Wang S**, **Xu B**, **Zhang Y**. 2013. Insect symbiont facilitates vector acquisition, retention, and transmission of plant virus. Sci Rep **3**:1367.

By generating genetically identical strains of the B. tabaci Q biotype where one strain harbors Hamiltonella and the other does not, the authors of this study show a strong effect of the symbiont Hamiltonella on the acquisition and transmission of TYLCV.

9. **Stores HH**. 1925. The Transmissiok of Streak Disease of Maize by the Leafhopper Balclutha Mbila Nalide. Ann Appl Biol **12**:422–439.

10. **Roosien BK**, **Gomulkiewicz R**, **Ingwell LL**, **Bosque-Pérez NA**, **Rajabaskar D**, **Eigenbrode SD**. 2013. Conditional vector preference aids the spread of plant pathogens: results from a model. Environ Entomol **42**:1299–1308.

11. **Ingwell LL**, **Eigenbrode SD**, **Bosque-Pérez NA**. 2012. Plant viruses alter insect behavior to enhance their spread. Sci Rep **2**:578.

12. **Mayer RT**, **Inbar M**, **McKenzie CL**, **Shatters R**, **Borowicz V**, **Albrecht U**, **Powell CA**, **Doostdar H**. 2002. Multitrophic interactions of the silverleaf whitefly, host plants, competing herbivores, and phytopathogens. Arch Insect Biochem Physiol **51**:151–169.

13. **Blua MJ**, **Perring TM**. 1992. Alatae production and population increase of aphid vectors on virus-infected host plants. Oecologia **92**:65–70.

14. **Gildow FE**. 1980. Increased Production of Alatae by Aphids Reared on Oats Infected with Barley Yellow Dwarf Virus. Ann Entomol Soc Am **73**:343–347.

15. **Castle SJ**, **Mowry TM**, **Berger PH**. 1998. Differential Settling by Myzus persicae (Homoptera: Aphididae) on Various Virus Infected Host Plants. Ann Entomol Soc Am **91**:661–667.

16. **Eigenbrode SD**, **Ding H**, **Shiel P**, **Berger PH**. 2002. Volatiles from potato plants infected with potato leafroll virus attract and arrest the virus vector, Myzus persicae (Homoptera: Aphididae). Proc Biol Sci **269**:455–460.

17. **Hodge S**, **Powell G**. 2008. Do plant viruses facilitate their aphid vectors by inducing symptoms that alter behavior and performance? Environ Entomol **37**:1573–1581.

18. **Mauck K**, **Bosque-Pérez NA**, **Eigenbrode SD**, **De Moraes CM**, **Mescher MC**. 2012. Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses. Funct Ecol **26**:1162–1175.

19. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. Proc Natl Acad Sci U S A **107**:3600–3605.

20. **Mauck KE**, **De Moraes CM**, **Mescher MC**. 2010. Effects of Cucumber mosaic virus infection on vector and non-vector herbivores of squash. Commun Integr Biol **3**:579–582.

21. **McMenemy LS**, **Hartley SE**, **MacFarlane SA**, **Karley AJ**, **Shepherd T**, **Johnson SN**. 2012. Raspberry viruses manipulate the behaviour of their insect vectors. Entomol Exp Appl **144**:56–68.

22. **Miller JW**, **Coon BF**. 1964. The Effect of Barley Yellow Dwarf Virus on the Biology of Its Vector the English Grain Aphid, Macrosiphum granarium. J Econ Entomol **57**:970–974.

23. **Castle SJ**, **Berger PH**. 1993. Rates of growth and increase of Myzus persicae on virus-infected potatoes according to type of virus-vector relationship. Entomol Exp Appl **69**:51–60.

24. **Luan J-B**, **Yao D-M**, **Zhang T**, **Walling LL**, **Yang M**, **Wang Y-J**, **Liu S-S**. 2013. Suppression of terpenoid synthesis in plants by a virus promotes its mutualism with vectors. Ecol Lett **16**:390–398.

25. **Matsuura S**, **Hoshino S**. 2009. Effect of tomato yellow leaf curl disease on reproduction of Bemisia tabaci Q biotype (Hemiptera: Aleyrodidae) on tomato [Lycopersicon esculentum] plants. Appl Entomol Zool Jpn.

26. **Maris PC**, **Joosten NN**, **Goldbach RW**, **Peters D**. 2004. Tomato spotted wilt virus Infection Improves Host Suitability for Its Vector Frankliniella occidentalis. Phytopathology **94**:706–711.

27. **Belliure B**, **Janssen A**, **Maris PC**, **Peters D**, **Sabelis MW**. 2005. Herbivore arthropods benefit from vectoring plant viruses. Ecol Lett **8**:70–79.

28. **Luan J-B**, **Ghanim M**, **Liu S-S**, **Czosnek H**. 2013. Silencing the ecdysone synthesis and signaling pathway genes disrupts nymphal development in the whitefly. Insect Biochem Mol Biol **43**:740–746.

29. **Blua MJ**, **Perring TM**, **Madore MA**. 1994. Plant virus-induced changes in aphid population development and temporal fluctuations in plant nutrients. J Chem Ecol **20**:691–707.

30. **Nanayakkara UN**, **Giguère MA**, **Pelletier Y**. 2012. Population Growth of Myzus persicae on Potato Plants Infected with Different Strains and Variants of Potato virus Y. Am J Potato Res **90**:297–300.

31. **Wosula EN**, **Davis JA**, **Clark CA**. 2013. Population dynamics of three aphid species (Hemiptera: Aphididae) on four Ipomoea spp. infected or noninfected with sweetpotato potyviruses. J Econ Entomol **106**:1566–1573.

32. **Ellsbury MM**, **Pratt RG**, **Knight WE**. 1985. Effects of Single and Combined Infection of Arrowleaf Clover with Bean Yellow Mosaic Virus and a Phytophthora sp. on Reproduction and Colonization by Pea Aphids (Homoptera: Aphididae). Environ Entomol **14**:356–359.

33. **Inoue T**, **Sakurai T**. 2006. Infection of *Tomato spotted wilt virus* (TSWV) shortens the life span of thelytokous *Thrips tabaci* (Thysanoptera: Thripidae). Appl Entomol Zool **41**:239–246.

34. **Donaldson JR**, **Gratton C**. 2007. Antagonistic effects of soybean viruses on soybean aphid performance. Environ Entomol **36**:918–925.

35. **Lei W**, **Liu D**, **Li P**, **Hou M**. 2014. Interactive Effects of Southern Rice Black-Streaked Dwarf Virus Infection of Host Plant and Vector on Performance of the Vector, Sogatella furcifera (Homoptera: Delphacidae). J Econ Entomol **107**:1721–1727.

36. **Kamali M**, **Xia A**, **Tu Z**, **Sharakhov IV**. 2012. A New Chromosomal Phylogeny Supports the Repeated Origin of Vectorial Capacity in Malaria Mosquitoes of the Anopheles gambiae Complex. PLOS Pathog **8**:e1002960.

37. **Czosnek H**, **Ghanim M**, **Morin S**, **Rubinstein G**, **Fridman V**, **Zeidan M**. 2001. Whiteflies: Vectors, and victims (?), of geminiviruses, p. 291–322. *In* Research, B-A in V (ed.), . Academic Press.

38. **Czosnek H**, **Ghanim M**. 2012. Back to Basics: Are Begomoviruses Whitefly Pathogens? J Integr Agric **11**:225–234.

39. **Jiu M**, **Zhou X-P**, **Tong L**, **Xu J**, **Yang X**, **Wan F-H**, **Liu S-S**. 2007. Vector-Virus Mutualism Accelerates Population Increase of an Invasive Whitefly. PLoS ONE **2**:e182.

40. **Zhang T**, **Luan J-B**, **Qi J-F**, **Huang C-J**, **Li M**, **Zhou X-P**, **Liu S-S**. 2012. Begomovirus–whitefly mutualism is achieved through repression of plant defences by a virus pathogenicity factor. Mol Ecol **21**:1294–1304.

41. **Frago E**, **Dicke M**, **Godfray HCJ**. 2012. Insect symbionts as hidden players in insect–plant interactions. Trends Ecol Evol **27**:705–711.

42. **Gray S**, **Gildow FE**. 2003. Luteovirus-aphid interactions. Annu Rev Phytopathol **41**:539–566.

43. **Gildow FE**. 1987. Virus—Membrane Interactions Involved in Circulative Transmission of Luteoviruses by Aphids, p. 93–120. *In* Harris, KF (ed.), Current Topics in Vector Research. Springer New York.

44. **Brault V**, **Tanguy S**, **Reinbold C**, **Le Trionnaire G**, **Arneodo J**, **Jaubert-Possamai S**, **Guernec G**, **Tagu D**. 2010. Transcriptomic analysis of intestinal genes following acquisition of pea enation mosaic virus by the pea aphid Acyrthosiphon pisum. J Gen Virol **91**:802–808.

45. **Baumann P**. 2005. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. Annu Rev Microbiol **59**:155–189.

46. **Moran NA**. 2002. The Ubiquitous and Varied Role of Infection in the Lives of Animals and Plants. Am Nat **160**:S1–S8.

47. **Hansen AK**, **Moran NA**. 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. Proc Natl Acad Sci **108**:2849–2854.

48. **Wilkinson TL**, **Douglas AE**. 1996. The impact of aposymbiosis on amino acid metabolism of pea aphids, p. 279–282. *In* Städler, E, Rowell-Rahier, M, Bauer, R (eds.), Proceedings of the 9th International Symposium on Insect-Plant Relationships. Springer Netherlands.

49. **Douglas AE**. 1998. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. Annu Rev Entomol **43**:17–37.

50. **Dennison NJ**, **Jupatanakul N**, **Dimopoulos G**. 2014. The mosquito microbiota influences vector competence for human pathogens. Curr Opin Insect Sci **3**:6–13.

This review discusses the interactions between insect vectors of human pathogens, their diverse microbiota and the pathogens they transmit. Pathogens might cause infection to both humans and to their insect vectors, depending on these tripartite interactions, and this can be useful to develop new management strategies.

51. **Poliakov A**, **Russell CW**, **Ponnala L**, **Hoops HJ**, **Sun Q**, **Douglas AE**, **van Wijk KJ**. 2011. Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. Mol Cell Proteomics MCP **10**:M110.007039.

52. **Shigenobu S**, **Stern DL**. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. Proc Biol Sci **280**:20121952.

53. **International Aphid Genomics Consortium**. 2010. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol **8**:e1000313.

54. **Douglas AE**, **Bouvaine S**, **Russell RR**. 2011. How the insect immune system interacts with an obligate symbiotic bacterium. Proc R Soc B Biol Sci **278**:333–338.

    This article showed that Buchnera cells induced an immune response via the IMD pathway in nonhost cells, i.e. Drosophila S2 cells, being eliminated from the nonhost cells in 1-2 days. The data supports the hypothesis that the absence of the IMD pathway and peptidoglycan recognition proteins, important in the recognition and defense against Gram-negative bacteria, is an adaptation of aphids to the symbiosis with Buchnera.

55. **Jaubert-Possamai S**, **Rispe C**, **Tanguy S**, **Gordon K**, **Walsh T**, **Edwards O**, **Tagu D**. 2010. Expansion of the miRNA pathway in the hemipteran insect Acyrthosiphon pisum. Mol Biol Evol **27**:979–987.

56. **Nyholm SV**, **Graf J**. 2012. Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. Nat Rev Microbiol **10**:815–827.

57. **Cilia M**, **Fish T**, **Yang X**, **McLaughlin M**, **Thannhauser TW**, **Gray S**. 2009. A comparison of protein extraction methods suitable for gel-based proteomic studies of aphid proteins. J Biomol Tech JBT **20**:201–215.

58. **van den Heuvel JFJM**, **Verbeek M**, **van der Wilk F**. 1994. Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by Myzus persicae. J Gen Virol **75**:2559–2565.

59. **Hogenhout SA**, **Ammar E-D**, **Whitfield AE**, **Redinbaugh MG**. 2008. Insect Vector Interactions with Persistently Transmitted Viruses. Annu Rev Phytopathol **46**:327–359.

60. **Kaplan IB**, **Lee L**, **Ripoll DR**, **Palukaitis P**, **Gildow F**, **Gray SM**. 2007. Point mutations in the potato leafroll virus major capsid protein alter virion stability and aphid transmission. J Gen Virol **88**:1821–1830.

61. **Soto MJ**, **Chen L-F**, **Seo Y-S**, **Gilbertson RL**. 2005. Identification of regions of the Beet mild curly top virus (family Geminiviridae) capsid protein involved in systemic infection, virion formation and leafhopper transmission. Virology **341**:257–270.

62. **Brault V**, **Bergdoll M**, **Mutterer J**, **Prasad V**, **Pfeffer S**, **Erdinger M**, **Richards KE**, **Ziegler-Graff V**. 2003. Effects of Point Mutations in the Major Capsid Protein of Beet Western Yellows Virus on Capsid Formation, Virus Accumulation, and Aphid Transmission. J Virol **77**:3247–3256.

63. **Reinbold C**, **Gildow FE**, **Herrbach E**, **Ziegler-Graff V**, **Gonçalves MC**, **van den Heuvel JFJM**, **Brault V**. 2001. Studies on the role of the minor capsid protein in transport of Beet western yellows virus through Myzus persicae. J Gen Virol **82**:1995–2007.

64. **Brault V**, **Mutterer J**, **Scheidecker D**, **Simonis MT**, **Herrbach E**, **Richards K**, **Ziegler-Graff V**. 2000. Effects of Point Mutations in the Readthrough Domain of the Beet Western Yellows Virus Minor Capsid Protein on Virus Accumulation In Planta and on Transmission by Aphids. J Virol **74**:1140–1148.

65. **Brault V**, **van den Heuvel JF**, **Verbeek M**, **Ziegler-Graff V**, **Reutenauer A**, **Herrbach E**, **Garaud JC**, **Guilley H**, **Richards K**, **Jonard G**. 1995. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. EMBO J **14**:650–659.

66. **Wen F**, **Lister RM**. 1991. Heterologous encapsidation in mixed infections among four isolates of barley yellow dwarf virus. J Gen Virol **72**:2217–2223.

67. **Chavez JD**, **Cilia M**, **Weisbrod CR**, **Ju H-J**, **Eng JK**, **Gray SM**, **Bruce JE**. 2012. Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus-plant interactions. J Proteome Res **11**:2968–2981.

68. **Watanabe S**, **Bressan A**. 2013. Tropism, compartmentalization and retention of banana bunchy top virus (Nanoviridae) in the aphid vector Pentalonia nigronervosa. J Gen Virol **94**:209–219.

69. **Filichkin SA**, **Brumfield S**, **Filichkin TP**, **Young MJ**. 1997. In vitro interactions of the aphid endosymbiotic SymL chaperonin with barley yellow dwarf virus. J Virol **71**:569–577.

70. **Bouvaine S**, **Boonham N**, **Douglas AE**. 2011. Interactions between a luteovirus and the GroEL chaperonin protein of the symbiotic bacterium Buchnera aphidicola of aphids. J Gen Virol **92**:1467–1474.

In this study, the involvement of Buchnera GroEL in facilitating virus transmission by aphids was called into question by presenting evidence that this bacterial protein was not found in the aphid hemolymph, gut or fat body, thus making it unlikely to interact with the virus. The authors also showed that the only possible binding site of GroEL to the virus would compete with other molecules of GroEL for the formation of the two GroEL rings, which is the functional form of the protein.

71. **Chaudhary R**, **Atamian HS**, **Shen Z**, **Briggs SP**, **Kaloshian I**. 2014. GroEL from the endosymbiont Buchnera aphidicola betrays the aphid by triggering plant defense. Proc Natl Acad Sci U S A **111**:8919–8924.

In this article, the authors claim to have detected Buchnera GroEL protein in aphid saliva and hypothesize that the Buchnera GroEL is recognized by plants as a bacterial molecular pattern and thus induces pattern-triggered immune responses. Aphids fed on Arabidopsis plants expressing the GroEL protein had reduced fecundity. Their explanation of these data is that Buchnera betrays its aphid host by inducing plant defenses that are detrimental to aphid reproduction.

72. **Vandermoten S**, **Harmel N**, **Mazzucchelli G**, **De Pauw E**, **Haubruge E**, **Francis F**. 2014. Comparative analyses of salivary proteins from three aphid species. Insect Mol Biol **23**:67–77.

73. **Chatellier J**, **Buckle AM**, **Fersht AR**. 1999. GroEL recognises sequential and non-sequential linear structural motifs compatible with extended β-strands and α-helices1. J Mol Biol **292**:163–172.

74. **Hinode D**, **Nakamura R**, **Grenier D**, **Mayrand D**. 1998. Cross-reactivity of specific antibodies directed to heat shock proteins from periodontopathogenic bacteria of human origin. Oral Microbiol Immunol **13**:55–58.

75. **Johansen HK**, **Nørgaard A**, **Andersen LP**, **Jensen P**, **Nielsen H**, **Høiby N**. 1995. Cross-reactive antigens shared by Pseudomonas aeruginosa, Helicobacter pylori, Campylobacter jejuni, and Haemophilus influenzae may cause false-positive titers of antibody to H. pylori. Clin Diagn Lab Immunol **2**:149–155.

76. **van den Heuvel JF**, **Bruyère A**, **Hogenhout SA**, **Ziegler-Graff V**, **Brault V**, **Verbeek M**, **van der Wilk F**, **Richards K**. 1997. The N-terminal region of the luteovirus readthrough domain determines virus binding to Buchnera GroEL and is essential for virus persistence in the aphid. J Virol **71**:7258–7265.

77. **Liu S**, **Sivakumar S**, **Wang Z**, **Bonning BC**, **Miller WA**. 2009. The readthrough domain of pea enation mosaic virus coat protein is not essential for virus stability in the hemolymph of the pea aphid. Arch Virol **154**:469–479.

    The authors of this study show that a mutant form of PEMV lacking the readthrough domain shows similar stability as wildtype PEMV in the aphid, providing evidence against a role for a stabilizing GroEL-readthrough protein interaction in the aphid during circulative transmission.

78. **Yang X**, **Thannhauser TW**, **Burrows M**, **Cox-Foster D**, **Gildow FE**, **Gray SM**. 2008. Coupling genetics and proteomics to identify aphid proteins associated with vector-specific transmission of polerovirus (luteoviridae). J Virol **82**:291–299.

79. **Cilia M**, **Howe K**, **Fish T**, **Smith D**, **Mahoney J**, **Tamborindeguy C**, **Burd J**, **Thannhauser TW**, **Gray S**. 2011. Biomarker discovery from the top down: Protein biomarkers for efficient virus transmission by insects (Homoptera: Aphididae) discovered by coupling genetics and 2-D DIGE. Proteomics **11**:2440–2458.

80. **Cilia M**, **Tamborindeguy C**, **Fish T**, **Howe K**, **Thannhauser TW**, **Gray S**. 2011. Genetics Coupled to Quantitative Intact Proteomics Links Heritable Aphid and Endosymbiont Protein Expression to Circulative Polerovirus Transmission. J Virol **85**:2148–2166.

    This study identified 50 aphid and Buchnera proteins differentially expressed between F2 genotypes of Schizaphis graminum from a cross between a vector and a non-vector of CYDV-RPV. In the F2 population, all efficient vector genotypes expressed the Buchnera protein isoforms similar to the competent parent, so the vectoring efficient phenotype was correlated to a specific genotype of Buchnera.

81. **Gray SM**, **Smith DM**, **Barbierri L**, **Burd J**. 2002. Virus Transmission Phenotype Is Correlated with Host Adaptation Among Genetically Diverse Populations of the Aphid Schizaphis graminum. Phytopathology **92**:970–975.

82. **Elzinga DA**, **De Vos M**, **Jander G**. 2014. Suppression of plant defenses by a Myzus persicae (green peach aphid) salivary effector protein. Mol Plant-Microbe Interact MPMI **27**:747–756.

In this study, Buchnera GroEL transiently expressed in Nicotiana tabacum, as well as stably expressed in transgenic Arabidopsis thaliana, reduced aphid reproduction. The authors' hypothesis is that the expression of GroEL activated antibacterial defense responses in the plants, impacting aphid fitness.

83. **Sloan DB**, **Moran NA**. 2012. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. Mol Biol Evol **29**:3781–3792.

84. **Jiang Z-F**, **Xia F**, **Johnson KW**, **Bartom E**, **Tuteja JH**, **Stevens R**, **Grossman RL**, **Brumin M**, **White KP**, **Ghanim M**. 2012. Genome Sequences of the Primary Endosymbiont "Candidatus Portiera aleyrodidarum" in the Whitefly Bemisia tabaci B and Q Biotypes. J Bacteriol **194**:6678–6679.

85. **Thao ML**, **Baumann P**. 2004. Evolutionary Relationships of Primary Prokaryotic Endosymbionts of Whiteflies and Their Hosts. Appl Environ Microbiol **70**:3401–3406.

86. **Gottlieb Y**, **Ghanim M**, **Chiel E**, **Gerling D**, **Portnoy V**, **Steinberg S**, **Tzuri G**, **Horowitz AR**, **Belausov E**, **Mozes-Daube N**, **Kontsedalov S**, **Gershon M**, **Gal S**, **Katzir N**, **Zchori-Fein E**. 2006. Identification and Localization of a Rickettsia sp. in Bemisia tabaci (Homoptera: Aleyrodidae). Appl Environ Microbiol **72**:3646–3652.

87. **Bing X-L**, **Ruan Y-M**, **Rao Q**, **Wang X-W**, **Liu S-S**. 2013. Diversity of secondary endosymbionts among different putative species of the whitefly Bemisia tabaci. Insect Sci **20**:194–206.

88. **Kontsedalov S**, **Zchori-Fein E**, **Chiel E**, **Gottlieb Y**, **Inbar M**, **Ghanim M**. 2008. The presence of Rickettsia is associated with increased susceptibility of Bemisia tabaci (Homoptera: Aleyrodidae) to insecticides. Pest Manag Sci **64**:789–792.

89. **Ghanim M**, **Kontsedalov S**. 2007. Gene expression in pyriproxyfen- resistant Bemisia tabaci Q biotype. Pest Manag Sci **63**:776–783.

90. **Russell JA**, **Moran NA**. 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proc R Soc Lond B Biol Sci **273**:603–610.

91. **Moran NA**, **Degnan PH**, **Santos SR**, **Dunbar HE**, **Ochman H**. 2005. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. Proc Natl Acad Sci U S A **102**:16919–16926.

92. **Oliver KM**, **Russell JA**, **Moran NA**, **Hunter MS**. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc Natl Acad Sci **100**:1803–1807.

93. **Scarborough CL**, **Ferrari J**, **Godfray HCJ**. 2005. Aphid Protected from Pathogen by Endosymbiont. Science **310**:1781–1781.

94. **Xue X**, **Li S-J**, **Ahmed MZ**, **Barro PJD**, **Ren S-X**, **Qiu B-L**. 2012. Inactivation of Wolbachia Reveals Its Biological Roles in Whitefly Host. PLOS ONE **7**:e48148.

95. **Sakurai M**, **Koga R**, **Tsuchida T**, **Meng X-Y**, **Fukatsu T**. 2005. Rickettsia Symbiont in the Pea Aphid Acyrthosiphon pisum: Novel Cellular Tropism, Effect on Host Fitness, and Interaction with the Essential Symbiont Buchnera. Appl Environ Microbiol **71**:4069–4075.

96. **De Barro PJ**, **Liu S-S**, **Boykin LM**, **Dinsdale AB**. 2011. Bemisia tabaci: a statement of species status. Annu Rev Entomol **56**:1–19.

97. **Gueguen G**, **Vavre F**, **Gnankine O**, **Peterschmitt M**, **Charif D**, **Chiel E**, **Gottlieb Y**, **Ghanim M**, **Zchori-Fein E**, **Fleury F**. 2010. Endosymbiont metacommunities, mtDNA diversity and the evolution of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. Mol Ecol **19**:4365–4376.

98. **Skaljac M**, **Zanic K**, **Ban SG**, **Kontsedalov S**, **Ghanim M**. 2010. Co-infection and localization of secondary symbionts in two whitefly species. BMC Microbiol **10**:142.

99. **Marubayashi JM**, **Kliot A**, **Yuki VA**, **Rezende JAM**, **Krause-Sakate R**, **Pavan MA**, **Ghanim M**. 2014. Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. PLOS ONE **9**:e108363.

100. **Pan H**, **Chu D**, **Liu B**, **Shi X**, **Guo L**, **Xie W**, **Carrière Y**, **Li X**, **Zhang Y**. 2013. Differential effects of an exotic plant virus on its two closely related vectors. Sci Rep **3**:2230.

101. **Morin S**, **Ghanim M**, **Zeidan M**, **Czosnek H**, **Verbeek M**, **van den Heuvel JFJM**. 1999. A GroEL Homologue from Endosymbiotic Bacteria of the

WhiteflyBemisia tabaciIs Implicated in the Circulative Transmission of Tomato Yellow Leaf Curl Virus. Virology **256**:75–84.

102. **Gottlieb Y**, **Lavy E**, **Kaufman M**, **Markovics A**, **Ghanim M**, **Aroch I**. 2012. A novel bacterial symbiont in the nematode Spirocerca lupi. BMC Microbiol **12**:133.

103. **Kontsedalov S**, **Gottlieb Y**, **Ishaaya I**, **Nauen R**, **Horowitz R**, **Ghanim M**. 2009. Toxicity of spiromesifen to the developmental stages of Bemisia tabaci biotype B. Pest Manag Sci **65**:5–13.

104. **Brumin M**, **Kontsedalov S**, **Ghanim M**. 2011. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. Insect Sci **18**:57–66.

105. **Caspi-Fluger A**, **Inbar M**, **Mozes-Daube N**, **Katzir N**, **Portnoy V**, **Belausov E**, **Hunter MS**, **Zchori-Fein E**. 2012. Horizontal transmission of the insect symbiont Rickettsia is plant-mediated. Proc Biol Sci **279**:1791–1796.

106. **Himler AG**, **Adachi-Hagimori T**, **Bergen JE**, **Kozuch A**, **Kelly SE**, **Tabashnik BE**, **Chiel E**, **Duckworth VE**, **Dennehy TJ**, **Zchori-Fein E**, **Hunter MS**. 2011. Rapid Spread of a Bacterial Symbiont in an Invasive Whitefly Is Driven by Fitness Benefits and Female Bias. Science **332**:254–256.

107. **Su Q**, **Oliver KM**, **Xie W**, **Wu Q**, **Wang S**, **Zhang Y**. 2015. The whitefly-associated facultative symbiont Hamiltonella defensa suppresses induced plant defences in tomato. Funct Ecol **29**:1007–1018.

This paper shows that the whitefly symbiont Hamiltonella imparts fitness benefits to the insect by reducing plant defense pathways against both herbivory and microbes. Such cross-talk between the pathways is an intriguing insight into how insects and their symbionts may have been selected to collaborate in the molecular arms race against the plant and how the plant was selected to respond against the attack on two fronts.

Chapter 4

A circulative plant virus modulates the aphid anti-viral immune system and the aphid-*Buchnera* symbiosis

Unpublished work, in prep for submission to BMC Genomics

Patricia V. Pinheiro[1,2,3], Yi Xu[4], Yi Zheng[2], Ana Rita Rebelo[2], Somayeh Fattahalhosseini[2], Rogerio Santos Dos Silva[2], Yimin Xu[2], James Giovannoni[2], Zhangjun Fei[2], Stewart Gray[4,6], and Michelle Cilia[2,4,6#]

[1] Department of Entomology, Cornell University, Ithaca, NY 14853
[2] Boyce Thompson Institute for Plant Research, Ithaca, NY 14853
[3] Embrapa Rice and Beans, Santo Antônio de Goiás, 75375-000, Brazil
[4] Section of Plant Pathology and Plant-Microbe Biology, School of Integrated Plant Science, Cornell University, Ithaca, NY 14853
[5] Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA
[6] USDA Agricultural Research Service, Emerging Pest and Pathogen Research Unit, Ithaca, NY 14853
#Address correspondence to Michelle Cilia, mlc@cornell.edu, michelle.cilia@ars.usda.gov

ABSTRACT

Plant viruses in the *Luteoviridae* are transmitted exclusively by aphids in a persistent, circulative and non-propagative manner. Although it is well accepted that viruses in the Luteoviridae do not replicate in their vectors, how these viruses remain quiescent in aphids is unknown. We hypothesized that small interfering RNA (siRNA) targeting *Potato leafroll virus* (PLRV) are produced in viruliferous *Myzus persicae* to suppress plant virus replication and we used next generation sequencing to test this hypothesis. However, no significant number of PLRV related siRNAs was detected in PLRV-viruliferous aphids; in support of the idea of the lack of luteovirid replication in their aphid vectors. Serendipitously, we discovered that PLRV-infected plants induce changes in the aphid's antiviral response against an aphid virus, *Myzus persicae Densovirus* (MpDNV), as well as in post-transcriptional gene regulation of aphids, and in the molecular interplay with the aphid's primary, obligate endosymbiont *Buchnera aphidicola.* A significant number of MpDNV-derived siRNA was generated in all aphid samples. Interestingly, aphids fed on PLRV-infected plants, but not on any of the other treatments, including PVY-infected plants, generated an abundance of unusually long siRNA against MpDNV and at the same time, a reduced number of 22-mer siRNAs, which is the common size range of siRNA produced by Dcr-2 in insects. Additionally, PLRV-viruliferous aphids had higher titer of MpDNV, suggesting that feeding on a plant infected with a circulative virus impairs the siRNA antiviral pathway in aphids. MpDNV was detected in a FISH assay in the nuclei and cytoplasm of aphid cells through the whole alimentary canal in both PLRV-

viruliferous and non-viruliferous *M. persicae*. A similar pattern of siRNA size distribution was observed when analyzing the siRNA aligned to the aphid genome (Mp-siRNA). PLRV-viruliferous aphids produced significantly less 22nt Mp-siRNA, compared to the other treatments, and an abundance of 26-27 nt Mp-siRNA, another unusual size of siRNA. These results show that a plant or viral component of the PLRV infected plant drives a change in post-transcriptional gene regulation via the aphid's siRNA pathway. Aphids also generated *B. aphidicola* tRNA-derived sRNAs, and the profile of these sRNA was unique in PLRV-fed aphids, suggesting that PLRV also alters the aphid-endosymbiont relationship. Our results also show that although PLRV-infected plants activate their siRNA pathway as an antiviral defense to PLRV, this pathway is not involved in the nonhost resistance mechanism in turnip to PLRV.

BACKGROUND

Aphids are the most important and wide spread vectors of plant viruses [1, 2]. Together, aphids and plant viruses cause significant crop yield losses around the world. Since there is no cure for virus infection in plants, plant host resistance to viruses and insecticide application to reduce aphid populations are the most common methods of plant virus management. However, for certain species of viruses and aphids, there are few options of commercially available resistant cultivars and of insecticides that effectively control aphid populations, due to the selection for aphid populations that are resistant to insecticides. Understanding

how aphids and plant viruses interact and how they have coevolved is the first step to disrupting their interaction, as a new and promising approach to control plant virus spread.

Transmission of plant viruses by aphids involves a combination of biological players that have co-evolved: the plant host, the aphid vector, the plant virus and the aphid obligate endosymbiont, *Buchnera aphidicola*. In this tug-of-war, plant hosts activate their immune defenses against aphids and plant viruses, and the latter two use a myriad of strategies to overcome, counteract or skip the host plant defenses. It has been suggested that insect vectors and the viruses they transmit might collaborate in fighting or avoiding host plant defenses [3–6]. The obligate endosymbiont of insect vectors might also play a role in virus transmission, either directly or indirectly [7].

Plant viruses from the family *Luteoviridae*, called herein as Luteovirids, are transmitted exclusively by aphids in a circulative mode of transmission (For a recent review, please see [8]) and they are retained in the aphid vector for its entire life, so they are classified as persistently transmitted viruses. Although controversial in the past [9–12], Luteovirids are currently considered as non-propagative viruses, which means that they do not replicate in the aphid vector [8]. Instead, aphids are considered as a passive vehicle for virus transmission and have been shown to benefit from feeding on plants infected with viruses they transmit [4, 6, 13]. Luteovirids are kept at low and harmless levels in aphid tissues. These viruses are retained in vesicles during transport across aphid cells, which makes them unavailable for exposure to the cytoplasm, where replication is

possible [14]. However, if a minority of virus particles were exposed to the cytoplasm of either gut or accessory salivary gland cells, a low level of virus replication in aphid cells would be possible. It is pertinent to compare the circulatively transmitted luteovirids to non-persistently transmitted viruses when considering the molecular interactions with the vector. In contrast to persistent circulative viruses, nonpersistently transmitted viruses, such as the Potyviridae *Potato Virus Y* (PVY), do not need to circulate through the insect vector tissues to be transmitted. Upon ingestion of nonpersistently transmitted viruses by an aphid, these viruses stay bound to the aphid stylet for a short period, from where they can be transmitted to a new host plant. Therefore, replication of nonpersistently transmitted viruses in their aphid vectors is not expected and the level of cellular interaction among aphids and these viruses is reduced, compared to circulative viruses.

Recently, “the vector manipulation hypothesis” has been proposed to explain the relationship between insect vectors and the viruses they transmit, borrowing the term from the animal vector field. The hypothesis predicts that a virus will promote its plant-to-plant spread by influencing the plant host selection and feeding behavior of the insect vector [15, 16], for example, by altering the host plant to produce and release volatiles that attract aphids [17–20]. Both persistent and nonpersistent viruses alter the host plant to make it more attractive to insect vectors, but they have different strategies to promote their plant to plant spread [16, 21]. Persistently transmitted viruses alter the host plant to enhance plant quality, so that the insect vector will have a long

term feeding on the infected plant [16–18, 21], which will potentially make aphids acquire more virions and become more viruliferous. In contrast, it has been suggested that nonpersistently transmitted viruses influence the plant phenotype to make it a reduced quality diet, then promoting a rapid insect vector dispersal [19, 20]. In both strategies, upon virus acquisition, insect vectors become more attracted to healthy plants, which completes the plant-to-plant spread cycle of the virus.

The sequencing of the aphid genome revealed that aphids possess the genetic machinery for RNA interference [22]. Viruses that replicate in insect tissues produce small interfering RNAs (siRNA) that can be measured using RNA sequencing technology (RNA-seq). There is evidence that the titer of Luteovirids in the aphid vector is reduced over time when aphids are transferred from an infected plant to a plant where the virus cannot replicate [9, 11, 23]., suggesting that Luteovirids are non-propagative in their aphid vectors. However, some of these authors suggest that the “possibility that PLRV replicates in *M. persicae* to a small extent cannot be completely excluded” [23]. Aphids remain viruliferous with persistently transmitted viruses, such as Luteovirids, for their entire life, even after transferred to a plant where the virus cannot replicate. We hypothesized that aphids shut down the replication of Luteovirids by using an antiviral mechanism, such as the siRNA pathway. To test our hypothesis, we used RNA-seq to test whether siRNAs are produced by *M. persicae* against PLRV. While we did not identify siRNAs produced in the aphid against PLRV in our study, our results showed

that plants infected with PLRV and PVY, viruses that are transmitted in two distinct modes, have profound differences in the antiviral immune response of aphids, post-transcriptional gene regulation, and the symbiosis with *Buchnera*.

METHODS

Aphids and viruses

Parthenogenic reproducing colonies of the green peach aphid *Myzus persicae* Sultz were maintained on caged physalis (*Physalis floridana*) at 20°C with an 18-hour photoperiod.

Three week old hairynight shade (*Solanum physalifolium,* HNS) plants were inoculated with cDNA clone of *Potato leafroll virus* (PLRV) wild type [24] to serve as source of virus for inoculation of potato plants by aphid feeding. Tobacco (*Nicotiana tabacum*) plants infected with Potato Virus Y (PVY) strain O [25] were used as inoculum for aphid transmission to the potato plants used in the experiments.

Aphid samples for Deep Sequencing of small RNAs

Aphids were allowed a three-day acquisition access period (AAP) on the following treatments prior to collecting aphids for RNA-seq: 1) PLRV-infected potato plants cv. Red Maria; 2) purified PLRV (50ug/ul) in 30% sucrose; 3) PVY-infected potato

plants cv. Goldrush; 4) mock inoculated potato plants (cv. Red Maria) and 5) 30% sucrose solution. After feeding on these treatments, aphids were transferred to a turnip plant cv Purple Top White Globe (nonhost for PLRV and PVY) for three days to clear the stylets and guts of nonacquired virus and any virus derived siRNAs that were generated in the plant and ingested along with plant sap. Since PVY is a noncirculative virus, the treatment with PVY-infected potato plants was included to test whether the transmission mechanism of the plant virus would have an effect on the aphid antiviral immune response, compared to PLRV, a circulative virus that is acquired by aphid cells.

Aphids were harvested from turnip plants and flash-frozen for siRNA isolation. For each treatment, three to four biological replicates were harvested, each containing a pool of approximately 700 aphids (50-80mg). To ensure that aphids acquired PLRV from the infected plants and the virus-laden diets, a subset of aphids were tested by RT-PCR, using primers that amplify a 660bp fragment, which included the Coat Protein (5'-CTAAAGATTTCCTCCCACGTGCG-3') and (5'-GGAGTGGGTGTTGGTTGTGGGC-3'), as described [26].

Plant samples for Deep Sequencing of small RNAs

Plant tissue samples were collected in the experiments described above from potato and turnip plants fed to PLRV-viruliferous aphids for siRNA isolation and sequencing. Potato samples were collected from plants that were systemically infected with PLRV, three weeks after inoculation and from leaves that were locally

infected, three days after inoculation. Turnip samples were collected three days after PLRV inoculation by aphids.

SiRNA isolation, library construction and sRNA sequencing

Small RNA were isolated from whole aphids and plant leaves using the mirPremier microRNA isolation kit (Sigma-Aldrich, SNC50). Small RNA libraries were constructed from 50-100 ng of sRNA, as described [27], with some modifications. A commercial small RNA 3' linker was used for adapter ligation (5'rApp-CTGTAGGCACCATCAAT-Amine 3') (New England Labs, S1315S). The reverse transcription primer 5'Amine-GACGTGTGCTCTTCCGATCT ATTGATGGTGCCTACA*G 3'was used to hybridize to the excess 3' adapter and forms the single stranded DNA adapter into a double-stranded DNA molecule. Three to four individual sRNA libraries were prepared for each aphid and plant treatment from purified sRNA using unique barcoded-adapters, which were pooled in four lanes and sequenced on an Illumina HiSeq2500 instruments, operating in "High Output Mode".

SiRNA Data Analysis

sRNA deep sequencing data were processed to remove sequencing adapters, low quality reads, and short reads (< 15 nt) using the sRNA clean script provided by VirusDetect (http://bioinfo.bti.cornell.edu/tool/VirusDetect/). The remaining reads

were aligned to ribosomal RNA database [28] using Bowtie [29] and the mapped reads were removed. The highly expressed sRNAs (>10RPM) were mapped to the reference genomes using Bowtie [29] allowing no mismatch. Reference genomes: *Potato leafroll virus* (NC_001747 and KC456053), *Buchnera aphidicola* F009 strain from *Myzus persicae* (CP002703), *Myzus persicae Densovirus* (AY148187), *Potato Virus Y* (EF026074), *Myzus persicae* G006 (draft genome available at aphidbase.org). The mapping depth of *Myzus persicae Densovirus* is generated using SAMtools [30]. These cleaned reads were also aligned to the mature tRNAs of *Buchnera aphidicola* using Bowtie. To look for miRNA candidates, the mapped loci and 200 bp flanking sequences on each side were extracted and then folded in silico using the RNAfold program [31]. Resulting folded structures were checked with miRcheck [32] to identify conserved miRNAs candidates.

In light of the recent finding that some *Myzus persicae Densovirus* sequences are integrated into the *M. persicae* genome [33], we checked whether the most abundant MpDNV-derived siRNA reads in our dataset localized to sites of integration in the aphid genome. Sequence of integrated MpDNV was obtained by aligning the *M. persicae* draft genome (clone G006) to the genome of MpDNV (AY148187), with a cut-off e-value of 10^{-5}.

Myzus persicae Densovirus **titer in PLRV viruliferous and non-viruliferous aphids**

Myzus persicae Densovirus (MpDNV) titers were quantified in aphids to determine

whether the presence of PLRV in aphids would impact the titer of MpDNV, using two approaches: 1) quantitative PCR (qPCR) and 2) droplet digital PCR (ddPCR).

In the first experiment, aphids were fed for three days on potato plants (cv. Red Maria) that were either inoculated with PLRV, mock inoculated or not inoculated (healthy). Pools of 10 aphids/replicate (4 replicates/treatment) were collected for DNA extraction using the DNeasy Blood and Tissue kit (Qiagen). qPCR reactions were prepared using 20ng/ul of DNA and qPCR primers specific to the MpDNV strain in our aphid colony, designed in an intergenic region from location 2718-2903 in the MpDNV genome (5'-TGACAATGGGTATATTCATTGACCT-3' and 5'-ATCGTGCGTCAAAAGAAACCCT-3'). A purified PCR product of the same region in the MpDNV genome was quantified by Nanodrop and used as a standard for absolute quantification of MpDNV in the aphid samples. The concentration of MpDNV was compared between treatments using a One-way Anova and means were compared by the t test.

In the second experiment, ddPCR was used to quantify the titer of MpDNV in single aphids (20 replicates/treatment) that were fed for three days on PLRV-infected HNS plants and on PLRV-free HNS plants (control), and then were transferred to turnip plants for gut clearing. For DNA extraction, a single (whole) aphid was homogenized using a micro pestle in 30 uL of squish buffer (per 10 mL consists of 9.8 mL dH2O, 100 uL 1M Tris pH 8.0, 20 uL 0.5M EDTA and 50 uL 5M NaCl) containing 0.6 ul of 10 mg/mL of Proteinase K. The insect homogenate was then incubated at 37°C for 30 minutes (Proteinase K digestion) followed by 95°C for 2 minutes (Proteinase K heat inactivation). After Proteinase K treatment the samples

were centrifuged for 7 min at 14k rpm. The supernatant was then transferred into a new tube and stored at -20°C until further use. A droplet digital PCR assay was developed for MpDNV using the QX100 droplet digital PCR system (Bio-Rad). The ddPCR reaction for MpDNV consisted of 10 uL of 2X ddPCR Evagreen SuperMix (Bio-Rad), 1 uL of each 10 uM MpDNV primers (same primers used in the previous experiment), 7 uL of dH2O and 2uL of DNA diluted at 1:800 in a final volume of a 20 uL reaction. A cartridge holder containing 20 uL of the ddPCR reaction and 70 uL of droplet generator oil for Evagreen (Bio-Rad) was placed into the QX100 droplet generator (Bio-Rad) where droplets were generated. Droplets were then transferred to a 96-well plate (Eppendorf) and the plate was sealed with an easy pierce foil seal (Bio-Rad). PCR amplification was carried out on the Applied Biosystems 2720 Thermocycler. The thermocycling conditions started at 95°C for 5 min, followed by 40 cycles of 95°C for 30 sec and 60C for 1 min, 1 cycle at 4°C for 5 min, 1 cycle at 90°C for 5 min and ending at 12°C. Following amplification, the plate was inserted into the droplet reader cassette and loaded into the droplet reader (Bio-Rad). The droplets were automatically read at a rate of 8 wells per 15 min. The ddPCR droplet data was analyzed using the QuantaSoft analysis software (Bio-Rad), which presents the target results as copies per uL of PCR mixture. The number of copies of MpDNV per ul was compared between treatments using the Kruskal-Wallis nonparametric test, since the Shapiro Wilk test for Normal distribution was significant ($p>0.0001$).

Localization of MpDNV in PLRV viruliferous and non-viruliferous aphids by

Fluorescence *In Situ* Hybridization (FISH)

Adults of *M. persicae* were fed either on PLRV-infected or healthy HNS plants for 24 hours prior to the Fluorescent in situ hybridization (FISH) assay. FISH was performed as previously described [34]. Briefly, specimens were fixed in Carnoy's fixative (chloroform-ethanol-glacial acetic acid, 6:3:1, vol/vol) for 5 minutes following gut dissection in 1x PBS and hybridized overnight in hybridization buffer (20 mM Tris-HCl, pH 8.0, 0.9 M NaCl, 0.01% [wt/vol] sodium dodecyl sulfate, 30% [vol/vol] formamide) containing 10 pmol fluorescent probes per ml. For specific targeting of PLRV and MpDNV viruses, PLRV (5′-TTTCCATTTCCCTTCCACAG-3′) [35] and designed DenR2 (5′-ATCGTGCGTCAAAAGAAACCCT-3′) DNA probes were used respectively. Nuclei were stained with 4′,6′-diamidino-2-phenylindole (DAPI; 0.1 mg ml−1). The stained guts were mounted in hybridization buffer and viewed under a Leica TCS-SP5 (Leica Microsystems Exton, PA USA) confocal microscope. At least 10 guts were viewed under the microscope to confirm reproducibility. Specificity of detection was confirmed using no-probe and PLRV-free controls.

RESULTS

RNA-seq summary

Three to four individual sRNA libraries were prepared for each aphid and plant treatment from purified sRNA using unique barcoded-adapters**. Each aphid library generated between 3 and 7 million reads, while the plant libraries**

generated between 1 and 2 million reads (Table 1). Aphid libraries generated reads varying from 15 to 40 nt in length, which we will call here as small RNA (sRNA) reads. We will call the small RNAs in the range of 21-24 nt as small interfering RNAs (siRNAs), because that is the common size range of siRNA generated by Dicer-2 like enzyme (Dcr-2) in **the siRNA pathways** in insects [36–38].

***M. persicae* does not generate siRNAs as an immune response against *Potato leafroll virus* (PLRV) but it does generate siRNAs against *Myzus persicae Densovirus* (MpDNV)**

The number of sRNA reads generated in aphids fed on PLRV-infected plants or PLRV-spiked diet treatments that aligned to the PLRV genome was very low (Table 1), indicating that aphids do not generate substantial amounts of siRNAs as an immune response to PLRV. Similarly, no significant number of sRNA reads mapped to PVY genome was identified in PVY-viruliferous aphids (Table 1). In contrast, a significant number of sRNA reads from all aphid samples aligned to the genome of the aphid virus *Myzus persicae Densovirus* (MpDNV) (GenBank Acc. No AY148187), indicating that the aphid siRNA machinery is functional as an immune defense against a virus that infects the aphid (MpDNV) but not against a plant virus that is transmitted by aphids in a non-propagative mode of transmission (PLRV and PVY).

PLRV-infected plant modulates the aphid immune system

All aphid samples generated significant number of sRNA reads that aligned to MpDNV genome (Table 1), ranging from 15 to 40 nt in length. The size distribution of the MpDNV-derived sRNA generated in aphids fed on PLRV-infected potato plants was unique, compared to the aphid samples, with an abundance of unusually long sRNAs of 33 to 40 nt generated only in these aphids (Fig. 1), suggesting that the circulative virus-infected plant influences the aphid immune system. For the other treatments, the most abundant size of sRNA reads aligned to MpDNV was 22 nt, which is in the range of the most common size of siRNA generated in insects. Interestingly, the number of 22 nt siRNAs was significantly lower in aphids fed on PLRV-infected plants (Fig. 1), suggesting that feeding on a circulative virus-infected plant impairs the aphid's siRNA pathway. Additionally, the titer of MpDNV was significantly higher in aphids fed on PLRV-infected plants, in comparison with aphids fed on healthy and mock inoculated plants, supporting the hypothesis that PLRV influences the aphid immune response to MpDNV (Fig. 2). The quantification of MpDNV in single aphids also showed that aphids fed on PLRV-infected hairynight shade (*Solanum physalifolium*, HNS) plants have significantly more copies of MpDNV than aphids fed on healthy HNS plants (357.5 ± 132.7 vs. 163.2 ± 96.8, respectively, Kruskal-Wallis test, p=0.02). Using a FISH assay, MpDNV was detected through the whole alimentary canal in both PLRV-viruliferous and non-viruliferous *M. persicae* (Figure 3, B–L). In the aphid stomach, MpDNV seems to localize in a non-punctuated pattern and scattered through the nuclei (Figure 3 B, C, E, H, I and K), as well as in the cytoplasm, possibly as virions (Figure 3 F, G, J and L). Although the FISH

assay can be used as a quantitative assay, our images did not show any consistent pattern that corroborates the difference in titer observed in the ddPCR and qPCR results. In some areas specially in the stomach, MpDNV and PLRV were overlapped and co-localized (Fig. G and H) but this was not observed as a rule. Our results are the first report of PLRV and MpDNV localization in the aphid gut using a FISH assay.

The total distribution of MpDNV-derived sRNA along the MpDNV genome was similar across treatments, with the majority of reads mapped to the C terminal of the negative strand of the virus genome (Fig. S1). Comparing only the 22nt and 34-38 nt sRNAs between aphid samples, we observed that the distribution along the MpDNV genome was also similar across treatments (Fig. 4), indicating that the preferential sites of cleavage of MpDNV is similar across treatments. To check whether the siRNAs obtained in our experiments were in the sites of integration of the MpDNV into the *M. persicae* genome [33], we aligned the two genomes and found eight regions with the highest similarity, based on a e-value cut-off of > 10^{-5} (Table S1). Five of these regions, matched to the *M. persicae* genome with 100% similarity (Table S1). As the MpDNV-derived sRNAs obtained in our dataset were evenly distributed along the MpDNV genome (Fig. S1), most of the siRNAs in our dataset represent a real antiviral defense of aphids to MpDNV infection. Analysis of the 33-38-mer sRNAs did not identify these unusual long sRNAs as miRNA candidates (not shown).

Little is known about the size of siRNAs produced in aphids or in insects in general as an antiviral immune defense. As a comparison of the size distribution

obtained for the MpDNV-derived siRNA, we analyzed the size distribution of the siRNA aligned to the *M. persicae* genome and found a pattern that was similar across treatments, except for a peak of 26 and 27 nt observed only in PLRV-viruliferous aphids (Fig. 5). For the most common size of siRNA, which is 22 nt, the same pattern of the reads aligned to MpDNV was observed for those aligned to *M. persicae* genome. In aphids fed on mock-inoculated potato, on sucrose diet without PLRV or on PVY-infected plants, the most abundant size of siRNA was 22 nt. Here again, PLRV-viruliferous aphids also generated less 22-nt sRNA, as observed for the MpDNV-derived sRNAs.

PLRV-derived siRNAs are produced in PLRV-infected potato plants but not in turnip

PLRV-infected potato samples presented a highly abundant number of siRNA reads that aligned to the PLRV genome, for both leaf tissues collected at 3 weeks post inoculation (WPI) and at 3 days post inoculation (DPI) (Table 1).

The number of sRNA reads that aligned to PLRV in potato 3DPI was lower than at 3WPI, but still significantly high, which indicates that three days was sufficient time for potato plants to mount their immune defense against PLRV infection. The majority of those reads were 21-22 nt (Fig. 6), which is in the range of the siRNA size reported for plants, including potato [39–41]. In contrast, the turnip plants used for gut clearing of PLRV-viruliferous aphids did not produce significant number of PLRV-derived sRNA (Table 1). Turnip is a nonhost of PLRV, which means that PLRV does not replicate in turnip. As a nonhost of PLRV, turnips were used to

clear the aphid guts of any virus particle not acquired by aphid cells, and in this process, PLRV-viruliferous aphids fed on turnip plants for three days. Three days was sufficient time for potato plants to mount an antiviral defense against PLRV, so if the siRNA was the mechanism mediating nonhost defense in turnip, it is likely that three days would be enough time for turnips to produce siRNAs against this virus. However, as turnip did not generate a substantial number of sRNAs that aligned to PLRV, we conclude that the mechanism for nonhost resistance to PLRV in turnips is not mediated by the siRNA pathway and that PLRV does not start the replication process in turnip. Other factors, such as the lack of receptors in the turnip cell membrane might drive the nonhost resistance mechanism.

Twelve percent of the Buchnera-derived siRNAs in aphids are tRNAs

About nine percent of the sRNAs from aphid samples aligned to *Buchnera aphidicola* genes (Table 1). Among these, 12% of the sRNAs, in average, are from aminoacyl-tRNAs (Table 1), Asparagine being the most abundant (Asn-tRNA) (Fig. 7). The proportion of sRNA aligned to each *Buchnera* aminoacyl-tRNAs varied across treatments (Fig. 7 and 8). Aphids fed on the sucrose diets lacking PLRV and on PVY-infected potato presented significant more Asn-tRNA (Fig. 7). In contrast, for most of the other aminoacyl-tRNAs with abundant number of reads, the proportion of sRNAs generated in aphids fed on a PLRV source or on healthy potato was higher, compared to aphids that fed on a sucrose diet lacking PLRV or on PVY-infected potato (Fig. 8). This result suggests that the amino acid

requirements of aphids that fed on an unbalanced sucrose diet are similar to that of aphids that fed on PVY-infected potato, which is in agreement with previous studies showing that aphids perceive a plant infected with a nonpersistent virus as a low quality diet [16, 19].

DISCUSSION

The “vector manipulation hypothesis” predicts that plant viruses promote their plant-to-plant spread by influencing the feeding behavior of their insect vectors [15, 16]. Our results show that aphids do not activate their siRNA pathway in response to PLRV acquisition, suggesting that this virus is not treated by the aphid immune system as a threat, or in other words, that aphids and Luteovirids are not in conflict.

Luteovirids are transmitted exclusively by aphids in a circulative and non-propagative manner, which means that they are acquired by aphid cells and circulate through aphid tissues, without replicating in the aphid vector [8, 10, 11]. Although some studies suggested that limited replication of Luteovirids in the aphid tissues may occur [12, 23, 42], there is no evidence that Luteovirids replicate in the aphid vector tissues as a pathogen, that is, at increasing rates, causing infection to aphids. On the contrary, there is evidence that the titer of *Potato leafroll virus* (PLRV) decrease over time in aphids when they are kept on a nonhost of the virus, such as turnip and Chinese cabbage plants [9, 11, 23]. The mechanism by which Luteovirids do not replicate in aphids is not fully understood [8], but it is

known that Luteovirids are acquired by aphid cells via a clathrin-mediated endocytosis mechanism and that they circulate in the aphid tissues internalized in membrane-bound vesicles [14, 43]. Electron microscopy studies showed that Luteovirids localize in the cytoplasm of plant phloem cells [44], which is the site of virus replication, but they have never been observed free in the cytoplasm of aphid cells [14, 43, 45–47], which makes it unlikely for them to replicate in aphids [43, 48]. Here, we show that the aphid siRNA pathway does not treat PLRV as a viral threat, providing additional evidence for the lack of Luteovirid propagation in aphid tissues.

Passing through aphid tissues unnoticed by the aphid immune system provides Luteovirids with an excellent vehicle for plant-to-plant spread, which is advantageous to these viruses. At the same time, aphids benefit from feeding on plants infected with viruses that cause no harm to aphids, providing aphids with fitness advantages, such as improved reproduction rates and survival [13]. In contrast, carrying stylet-borne viruses, such as *Potato Virus Y* (PVY) from the Potyviridae family, is not as beneficial to aphids as carrying circulative viruses, although it still represents a fitness advantage to aphids in comparison to feeding on healthy plants [13]. Stylet-borne viruses, or nonpersistent viruses, do not need to circulate through aphid tissues to be transmitted. Upon ingestion, nonpersistent viruses are retained on the end of the aphid stylet and there is no evidence that they enter aphid cells. Therefore, it is expected that the aphid antiviral immune responses will not recognize or respond to nonpersistent viruses. Our results show that aphids do not activate their antiviral immune pathway against these two plant

viruses that are non-propagative in aphids, but they do activate their antiviral defense in response to infection with an aphid virus, *Myzus persicae Densovirus* (MpDNV), which replicates in the aphid.

The major cellular antiviral defense in eukaryotes, including insects, is mediated by the small interference RNA (siRNA) pathway [38, 49]. Upon virus infection, double stranded RNA (dsRNA) is generated by RNA viruses during their genome replication or as a byproduct of DNA viruses convergent transcription by bidirectional promoters [50–53]. DsRNA molecules generated by viruses are recognized by pathogen-associated molecular pattern (PAMP) receptors in the cytoplasm, triggering the cleavage of dsRNA into small short interfering RNA duplexes (siRNA) by the RNase enzyme Dcr-2. The siRNA pathway has been shown to be functional [54–58]. The only aphid species with a sequenced genome available to date is the pea aphid *Acyrtosiphon* pisum [22], but our study also benefited from the draft genome of *M. persicae*, which is soon to be released (aphidbase.org). The immune system of the pea aphid differs from that of other insect species that had their genomes sequenced until now in many aspects. For example, the pea aphid lacks components of many of the common pathways activated in other insect species as an immune response against bacteria [22, 59], which has been suggested to be related to the obligate symbiosis with *B. aphidicola*. On the contrary, the pea aphid have homologs for all core genes of antiviral defense pathways, such as the Toll signaling, JAK-STAT and siRNA pathways [59]. The genes involved in the siRNA pathway, *dcr-2*, *ago-2* and *r2d2*, are present in single copies in the pea aphid genome [55]. Although the antiviral

defenses seem to be conserved and functional in aphids, little is known about the aphid immune responses against virus infection. Here we show that the antiviral immune defense of aphids is altered when feeding on a plant infected with a circulative virus transmitted by the aphid, as a molecular signature of the vector manipulation hypothesis for the system *M. persicae*-PLRV.

Aphids fed on PLRV-infected plants, but not on any of the other treatments, including PVY-infected plants, generated an abundance of unusually long siRNA against MpDNV and at the same time, a reduced number of 22-mer siRNAs, which is the common size range of siRNA produced by Dcr-2 in insects [36–38]. These data suggest that in aphids fed on PLRV-infected plants, the siRNA pathway is altered or that an alternative enzyme or antiviral pathway is activated. Alignment of the long sRNA generated in PLRV-viruliferous aphids shows that these long sRNAs may be immature siRNAs that were not processed into the functional 22-nt form or they may be a product of a different enzyme in another pathway. We also observed that the titer of MpDNV is higher in PLRV-viruliferous aphids than in aphids fed on healthy plants. This result suggests that the antiviral pathway in aphids fed on PLRV-infected plants is less intense compared to aphids fed on the other treatments, where the Dcr-2-siRNA pathway is fully active.

MpDNV is a single-stranded DNA virus classified in the *Parvoviridae* family [60], with a genome of approximate 5.7kb [61]. Only minor effects of MpDVN on the aphid reproduction and development have been reported [60]. Icosahedral particles of MpDNV were reported to localize in the cytoplasm of the aphid stomach cells but not in the posterior midgut and hindgut aphid cells [60]. Recently, genomic

sequences of *M. persicae* were shown to share partial homology with MpDNV and the integrated viral sequences are transcribed in *M. persicae,* generating amino acid sequences that share 33 to 51% identity with MpDVN proteins [33]. In our dataset, five regions of the MpDNV matched the *M. persicae* draft genome that we used as a reference with 100% similarity, but no bias was observed in the distribution of sRNA reads along the MpDNV genome. These regions with high similarity sequences were not overrepresented in our sRNA dataset, therefore they might represent an actual antiviral defense of aphids to MpDNV infection. Replication of MpDNV in *M. persicae* indicates that their relationship is a host-pathogen relationship, *sensu stricto*, unlike the relationship between aphids and Luteovirids.

Our data also provided interesting information on the nonhost immunity of turnip to PLRV. Virus infection is a species specific process that depends on the recognition of host receptors. Most plants are immune to a certain species of virus, as attempts to cause infection result in no detectable symptom or virus multiplication. This lack of interaction is one of the most intriguing plant-virus relationships and it is known as nonhost immunity [64]. Little is known about host range determination or nonhost immunity to viruses in plant hosts. Plants activate their siRNA pathway as an immune defense against infective viruses, generating 21 to 24-nt siRNA (recently reviewed in [65]. PLRV infects plants in the Solanacea family, such as potato (*Solanum tuberosum*), its primary host. However, PLRV does not infect/replicate in turnip (*Brassica rapa*), a plant from the Brassicaceae family. Turnip resistance to PLRV is an example of nonhost immunity, for which

the mechanism is unknown. To our knowledge, our results are the first report showing that nonhost immunity to PLRV in turnip is not a siRNA-based mechanism. Future work should focus on different plant-virus systems to elucidate whether the lack of siRNA response is a pattern in nonhost immunity.

Finally, we also report the generation of *B. aphidicola* tRNA-derived sRNAs in aphids. *Buchnera* is the aphid obligate endosymbiont, providing aphids with essential amino acids they are unable to synthesize *de novo* or to obtain from their diet [66, 67]. Essential amino acids are those that animals cannot synthesize *de novo* and are typically obtained from the diet. However, the phloem sap is a poor source of the essential amino acids for aphids [68, 69] (Arg, His, Ile, Leu, Lys, Met, Phe, Thr, Trp and Val) [70]. *Buchnera*'s genome has most of the genes for the synthesis of essential amino acids, but some genes in these pathways are missing, as well as most of the genes for the synthesis of nonessential amino acids [71]. Therefore, aphids also provide *Buchnera* with nonessential amino acids, including some that are used by the bacteria as a precursor in the biosynthesis of essential amino acids [22, 71, 72]. Unlike other insect species with a genome sequence available, the pea aphid lacks genes for the synthesis of Arginine, which is synthesized by *Buchnera* [72]. Differential expression of genes in amino acid biosynthesis pathways between bacteriocytes and other aphid tissues indicate complementarity between amino acid pathways encoded by the host and symbiont genomes [71]. Seven nonessential amino acids are not synthesized by *Buchnera* (Glu, Asp, Ser, Gln, Ala, Pro and Asn) [73] and their pathways were upregulated in the bacteriocyte compared to the body of the pea aphid, except for proline and

asparagine [71]. Asparagine-tRNA (tRNA15-Asn-GTT) was the most abundant *Buchnera* gene represented in the sRNA dataset from aphids in our study. SRNAs aligned to Asn-tRNA were significantly less abundant in aphids that acquired PLRV, compared to aphids that were fed on a sucrose diet lacking the balanced amino acid amounts required by aphids and to aphids fed on a PVY-infected plant, which is also perceived by aphids as a poor diet [18, 19]. Glutamine and asparagine are the most abundant amino acids in the pea aphid hemolymph [74, 75] and asparagine is also the most abundant amino acid in the phloem of host plants of the pea aphid [68].

As in other endosymbiont genomes, *Buchnera* has lost most of the genes for transcriptional regulation that are present in free-living relatives, such as *Escherichia coli* [73]. Transcription in *Buchnera* is stable, with no detectable differential expression of mRNA [76]. Other mechanisms for regulation of gene expression in *Buchnera* have been proposed, such as small RNAs and regulated protein stability [76]. We show that nine percent of the sRNA reads from *M. persicae* samples are from the *Buchnera* genome, with a remarkable number of reads aligned to tRNAs. Recently, tRNA fragments have been recognized to play regulatory roles [77], but the function of the Buchnera tRNA-derived sRNAs in aphids is unknown. It has been suggested that endosymbionts of Hemipteran vectors of plant viruses might directly or indirectly play a role in virus transmission by their insect host but the mechanism for how this may be achieved is a highly contested topic in the field. (For a recent review, please see [78]. We show here that in aphids viruliferous with PLRV, the profile of *Buchnera* tRNA-derived sRNAs

was different than in PLRV-free aphids. For example, Asn-tRNA-derived sRNAs were less abundant in PLRV-viruliferous aphids, while the nonessential amino acids Gln, Leu and Ser were more abundant in PLRV-viruliferous aphids, suggesting that carrying PLRV alters the regulation of amino acid synthesis in the aphid-*Buchnera* relationship, and this might be a cause, consequence, or totally unrelated to the difference observed in the antiviral response of aphids to MpDNV, when viruliferous with PLRV. Altogether, our data shows that the functional boundaries of vector, host and pathogen are blurred in the interactions among these organisms.

CONCLUSIONS

- Feeding on a PLRV-infected plant affects the aphid antiviral response to an insect infecting virus (MpDNV).
- Our data support no luteovirid replication in aphid tissues – aphids are not a host of this virus.
- siRNA is not the mechanism involved in nonhost resistance/immunity in turnip.
- Feeding on a PLRV-infected plant affects the regulation of amino acid synthesis in the aphid-Buchnera relationship.

Funding

Funding for the present research was graciously provided by NSF grant 1354309

(to MC), and an Embrapa Graduate Student Fellowship (to PVP). DNA sequence data were downloaded from AphidBase, http://www.aphidbase.com/aphidbase/. Funding for *Myzus persicae* clone G006 genomic sequencing was provided by USDA-NIFA award 2010-65105-20558.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Jason Ingram (USDA ARS) for support with pesticide applications.

REFERENCES

1. Nault LR: **Arthropod transmission of plant viruses: A new synthesis**. *Ann Entomol Soc Am* 1997, **90**:521–541.

2. Ng JCK, Falk BW: **Virus-vector interactions mediating nonpersistent and semipersistent transmission of plant viruses**. *Annu Rev Phytopathol* 2006, **44**:183–212.

3. Hodge S, Powell G: **Do plant viruses facilitate their aphid vectors by inducing symptoms that alter behavior and performance?** *Environ Entomol* 2008, **37**:1573–1581.

4. Hodge S, Powell G: **Conditional facilitation of an aphid vector, Acyrthosiphon pisum, by the plant pathogen, pea enation mosaic virus**. *J Insect Sci* 2010, **10**:155.

5. Jiu M, Zhou X-P, Tong L, Xu J, Yang X, Wan F-H, Liu S-S: **Vector-virus mutualism accelerates population increase of an invasive whitefly**. *PLoS ONE* 2007, **2**:e182.

6. Kersch-Becker MF, Thaler JS: **Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores**. *Oecologia* 2014, **174**:883–892.

7. Pinheiro PV, Kliot A, Ghanim M, Cilia M: **Is there a role for symbiotic bacteria in plant virus transmission by insects?** *Curr Opin Insect Sci* 2015, **8**:69–78. [*Ecology * Parasites/Parasitoids/Biological Control*]

8. Gray S, Cilia M, Ghanim M: **Circulative, "nonpropagative" virus transmission: an orchestra of virus-, insect-, and plant-derived instruments**. *Adv Virus Res* 2014, **89**:141–199.

9. Day M: **The Mechanism of the Transmission of Potato Leaf Roll Virus by Aphids**. *Aust J Biol Sci* 1955, **8**:498–513.

10. Eskandari F, Sylvester E, Richardson J: **Evidence for Lack of Propagation of Potato Leaf Roll Virus in Its S Hid Vector, Myzus-Persicae**. *Phytopathology* 1979, **69**:45–47.

11. Harrison BD: **Studies on the behavior of potato leaf roll and other viruses in the body of their aphid vector Myzus persicae (Sulz.)**. *Virology* 1958, **6**:265–277.

12. Weidemann H-L: **Zur Vermehrung des Kartoffelblattrollvirus in der Blattlaus Myzus persicae (Sulz.)**. *Z Für Angew Entomol* 1982, **94**:321–330.

13. Castle SJ, Berger PH: **Rates of growth and increase of Myzus persicae on virus-infected potatoes according to type of virus-vector relationship**. *Entomol Exp Appl* 1993, **69**:51–60.

14. Gildow FE: **Virus—membrane interactions involved in circulative transmission of luteoviruses by aphids**. In *Current Topics in Vector Research*. Edited by Harris KF. Springer New York; 1987:93–120. [*Current Topics in Vector Research*, vol. 4]

15. Ingwell LL, Eigenbrode SD, Bosque-Pérez NA: **Plant viruses alter insect behavior to enhance their spread**. *Sci Rep* 2012, **2**:578.

16. Mauck K, Bosque-Pérez NA, Eigenbrode SD, De Moraes CM, Mescher MC: **Transmission mechanisms shape pathogen effects on host–vector interactions: evidence from plant viruses**. *Funct Ecol* 2012, **26**:1162–1175.

17. Eigenbrode SD, Ding H, Shiel P, Berger PH: **Volatiles from potato plants infected with potato leafroll virus attract and arrest the virus vector,**

Myzus persicae (Homoptera: Aphididae). *Proc Biol Sci* 2002, **269**:455–460.

18. Alvarez AE, Garzo E, Verbeek M, Vosman B, Dicke M, Tjallingii WF: **Infection of potato plants with potato leafroll virus changes attraction and feeding behaviour of Myzus persicae**. *Entomol Exp Appl* 2007, **125**:135–144.

19. Mauck KE, De Moraes CM, Mescher MC: **Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts**. *Proc Natl Acad Sci U S A* 2010, **107**:3600–3605.

20. McMenemy LS, Hartley SE, MacFarlane SA, Karley AJ, Shepherd T, Johnson SN: **Raspberry viruses manipulate the behaviour of their insect vectors**. *Entomol Exp Appl* 2012, **144**:56–68.

21. Castle SJ, Mowry TM, Berger PH: **Differential settling by Myzus persicae (Homoptera: Aphididae) on various virus infected host plants**. *Ann Entomol Soc Am* 1998, **91**:661–667.

22. International Aphid Genomics Consortium: **Genome sequence of the pea aphid Acyrthosiphon pisum**. *PLoS Biol* 2010, **8**:e1000313.

23. Tamada T, Harrison BD: **Quantitative studies on the uptake and retention of Potato leafroll virus by aphids in laboratory and field conditions**. *Ann Appl Biol* 1981, **98**:261–276.

24. Franco-Lara LF, McGeachy KD, Commandeur U, Martin RR, Mayo MA, Barker H: **Transformation of tobacco and potato with cDNA encoding the full-length genome of potato leafroll virus: evidence for a novel virus distribution and host effects on virus multiplication**. *J Gen Virol* 1999, **80 (Pt 11)**:2813–2822.

25. Karasev AV, Hu X, Brown CJ, Kerlan C, Nikolaeva OV, Crosslin JM, Gray SM: **Genetic diversity of the ordinary strain of Potato virus Y (PVY) and origin of recombinant PVY strains**. *Phytopathology* 2011, **101**:778–785.

26. Peter KA, Liang D, Palukaitis P, Gray SM: **Small deletions in the potato leafroll virus readthrough protein affect particle morphology, aphid transmission, virus movement and accumulation**. *J Gen Virol* 2008, **89**(Pt 8):2037–2045.

27. Chen Y-R, Zheng Y, Liu B, Zhong S, Giovannoni J, Fei Z: **A cost-effective method for Illumina small RNA-Seq library preparation using T4 RNA ligase 1 adenylated adapters**. *Plant Methods* 2012, **8**:41.

28. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.** *Nucleic Acids Res* 2013, **41**(Database issue):D590-6.

29. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.

30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinforma Oxf Engl* 2009, **25**:2078–9.

31. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: **ViennaRNA Package 2.0**. *Algorithms Mol Biol AMB* 2011, **6**:26.

32. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: **A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana**. *Genes Dev* 2006, **20**:3407–3425.

33. Clavijo G, van Munster M, Monsion B, Bochet N, Brault V: **Transcription of densovirus endogenous sequences in Myzus persicae genome**. *J Gen Virol* 2016.

34. Kliot A, Cilia M, Czosnek H, Ghanim M: **Implication of the bacterial endosymbiont Rickettsia spp. in interactions of the whitefly Bemisia tabaci with tomato yellow leaf curl virus**. *J Virol* 2014, **88**:5652–5660.

35. Ghanim M, Brumin M, Popovski S: **A simple, rapid and inexpensive method for localization of Tomato yellow leaf curl virus and Potato leafroll virus in plant and insect vectors**. *J Virol Methods* 2009, **159**:311–314.

36. Aliyari R, Wu Q, Li H-W, Wang X-H, Li F, Green LD, Han CS, Li W-X, Ding S-W: **Mechanism of Induction and Suppression of Antiviral Immunity Directed by Virus-Derived Small RNAs in Drosophila**. *Cell Host Microbe* 2008, **4**:387–397.

37. Sabin LR, Zheng Q, Thekkat P, Yang J, Hannon GJ, Gregory BD, Tudor M, Cherry S: **Dicer-2 Processes Diverse Viral RNA Species**. *PLoS ONE* 2013, **8**:e55458.

38. Xu Y, Huang L, Fu S, Wu J, Zhou X: **Population Diversity of Rice Stripe Virus-Derived siRNAs in Three Different Hosts and RNAi-Based Antiviral Immunity in Laodelphgax striatellus**. *PLoS ONE* 2012, **7**:e46238.

39. Denis Kutnjak, Rupar M, Gutierrez-Aguirre I, Curk T, Kreuze JF, Ravnikar M: **Deep sequencing of virus derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscape of a plant virus population**. *J Virol* 2015:JVI.03685-14.

40. Hwang YT, Kalischuk M, Fusaro AF, Waterhouse PM, Kawchuk L: **Small RNA sequencing of Potato leafroll virus-infected plants reveals an additional subgenomic RNA encoding a sequence-specific RNA-binding protein**. *Virology* 2013, **438**:61–69.

41. Li R, Gao S, Hernandez AG, Wechter WP, Fei Z, Ling K-S: **Deep Sequencing of Small RNAs in Tomato for Virus and Viroid Identification and Strain Differentiation**. *PLoS ONE* 2012, **7**:e37127.

42. Stegwee D, Ponsen MB: **Multiplication of potato leaf roll virus in the aphid Myzus persicae (Sulz.).** *Entomol Exp Appl* 1958, **1**:291–300 pp.

43. Garret A, Kerlan C, Thomas D: **Ultrastructural study of acquisition and retention of potato leafroll luteovirus in the alimentary canal of its aphid vector, Myzus persicae Sulz**. *Arch Virol* 1996, **141**:1279–1292.

44. Shepardson S, Esau K, McCrum R: **Ultrastructure of potato leaf phloem infected with potato leafroll virus**. *Virology* 1980, **105**:379–392.

45. Garret A, Kerlan C, Thomas D: **Visualization of a luteovirus in the vector aphid's body by two gold immunolabelling techniques: a comparative study**. *J Virol Methods* 1991, **35**:93–107.

46. Garret A, Kerlan C, Thomas D: **The intestine is a site of passage for potato leafroll virus from the gut lumen into the haemocoel in the aphid vector, Myzus persicae Sulz**. *Arch Virol* 1993, **131**:377–392.

47. Reinbold C, Gildow FE, Herrbach E, Ziegler-Graff V, Gonçalves MC, van den Heuvel JFJM, Brault V: **Studies on the role of the minor capsid protein in transport of Beet western yellows virus through Myzus persicae**. *J Gen Virol* 2001, **82**:1995–2007.

48. Peiffer ML, Gildow FE, Gray SM: **Two distinct mechanisms regulate luteovirus transmission efficiency and specificity at the aphid salivary gland**. *J Gen Virol* 1997, **78 (Pt 3)**:495–503.

49. Brackney DE, Beane JE, Ebel GD: **RNAi Targeting of West Nile Virus in Mosquito Midguts Promotes Virus Diversification**. *PLoS Pathog* 2009, **5**:e1000502.

50. Kato H, Takeuchi O, Mikamo-Satoh E, Hirai R, Kawai T, Matsushita K, Hiiragi A, Dermody TS, Fujita T, Akira S: **Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid–inducible gene-I and melanoma differentiation–associated gene 5**. *J Exp Med* 2008, **205**:1601–1610.

51. Pfaller CK, Radeke MJ, Cattaneo R, Samuel CE: **Measles Virus C Protein Impairs Production of Defective Copyback Double-Stranded Viral RNA and Activation of Protein Kinase R**. *J Virol* 2014, **88**:456–468.

52. Son K-N, Liang Z, Lipton HL: **Double-Stranded RNA Is Detected by Immunofluorescence Analysis in RNA and DNA Virus Infections, Including Those by Negative-Stranded RNA Viruses**. *J Virol* 2015, **89**:9383–9392.

53. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR: **Double-Stranded RNA Is Produced by Positive-Strand RNA Viruses and DNA Viruses but Not in Detectable Amounts by Negative-Strand RNA Viruses**. *J Virol* 2006, **80**:5059–5064.

54. Jaubert-Possamai S, Le Trionnaire G, Bonhomme J, Christophides GK, Rispe C, Tagu D: **Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum**. *BMC Biotechnol* 2007, **7**:63.

55. Jaubert-Possamai S, Rispe C, Tanguy S, Gordon K, Walsh T, Edwards O, Tagu D: **Expansion of the miRNA pathway in the hemipteran insect Acyrthosiphon pisum**. *Mol Biol Evol* 2010, **27**:979–987.

56. Pitino M, Coleman AD, Maffei ME, Ridout CJ, Hogenhout SA: **Silencing of aphid genes by dsRNA feeding from plants**. *PLoS ONE* 2011, **6**.

57. Sapountzis P, Duport G, Balmand S, Gaget K, Jaubert-Possamai S, Febvay G, Charles H, Rahbé Y, Colella S, Calevro F: **New insight into the RNA interference response against cathepsin-L gene in the pea aphid, Acyrthosiphon pisum: molting or gut phenotypes specifically induced by injection or feeding treatments**. *Insect Biochem Mol Biol* 2014, **51**:20–32.

58. Whyard S, Singh AD, Wong S: **Ingested double-stranded RNAs can act as species-specific insecticides**. *Insect Biochem Mol Biol* 2009, **39**:824–832.

59. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldón T, Ghanim M, Heddi A, Kaloshian I, Latorre A, Moya A, Nakabachi A, Parker BJ, Pérez-Brocal V, Pignatelli M, Rahbé Y, Ramsey JS, Spragg CJ, Tamames J, Tamarit D, Tamborindeguy C, Vincent-

Monegat C, Vilcinskas A: **Immunity and other defenses in pea aphids, Acyrthosiphon pisum**. *Genome Biol* 2010, **11**:R21.

60. van Munster M, Dullemans AM, Verbeek M, van den Heuvel JFJM, Reinbold C, Brault V, Clérivet A, van der Wilk F: **Characterization of a new densovirus infecting the green peach aphid Myzus persicae**. *J Invertebr Pathol* 2003, **84**:6–14.

61. van Munster M, Dullemans AM, Verbeek M, van den Heuvel JFJM, Reinbold C, Brault V, Clérivet A, van der Wilk F: **A new virus infecting Myzus persicae has a genome organization similar to the species of the genus Densovirus**. *J Gen Virol* 2003, **84**(Pt 1):165–172.

62. Huang Y-JS, Higgs S, Horne KM, Vanlandingham DL: **Flavivirus-Mosquito Interactions**. *Viruses* 2014, **6**:4703–4730.

63. Pakkianathan BC, Kontsedalov S, Lebedev G, Mahadav A, Zeidan M, Czosnek H, Ghanim M: **Replication of Tomato Yellow Leaf Curl Virus in Its Whitefly Vector, Bemisia tabaci**. *J Virol* 2015, **89**:9791–9803.

64. Fraser RSS: **Host-Range Control and Non-Host Immunity to Viruses**. In *Mechanisms of Resistance to Plant Diseases*. Edited by Fraser RSS. Springer Netherlands; 1985:13–28. [*Advances in Agricultural Biotechnology*, vol. 17]

65. Zhang C, Wu Z, Li Y, Wu J: **Biogenesis, Function, and Applications of Virus-Derived Small RNAs in Plants**. *Front Microbiol* 2015, **6**:1237.

66. Baumann P: **Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects**. *Annu Rev Microbiol* 2005, **59**:155–189.

67. Buchner P: *Endosymbiosis of Animals with Plant Microorganisms*. Interscience Publishers; 1965.

68. Sandstrom J, Pettersson J: **Amino-Acid-Composition of Phloem Sap and the Relation to Intraspecific Variation in Pea Aphid (acyrthosiphon-Pisum) Performance**. *J Insect Physiol* 1994, **40**:947–955.

69. Sandström JP, Moran NA: **Amino acid budgets in three aphid species using the same host plant**. *Physiol Entomol* 2001, **26**:202–211.

70. Mittler TE: **Dietary Amino Acid Requirements of the Aphid Myzus persicae Affected by Antibiotic Uptake**. *J Nutr* 1971, **101**:1023–1028.

71. Hansen AK, Moran NA: **Aphid genome expression reveals host–symbiont cooperation in the production of amino acids**. *Proc Natl Acad Sci* 2011, **108**:2849–2854.

72. Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE: **Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola**. *Insect Mol Biol* 2010, **19 Suppl 2**:249–258.

73. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS**. *Nature* 2000, **407**:81–86.

74. Sasaki T, Ishikawa H: **Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, Acyrthosiphon pisum**. *J Insect Physiol* 1995, **41**:41–46.

75. Sasaki T, Ishikawa H: **Nitrogen Recycling in the Endosymbiotic System of the Pea Aphid, Acyrthosiphon pisum**. *Zoolog Sci* 1993, **10**:779–785.

76. Hansen AK, Degnan PH: **Widespread expression of conserved small RNAs in small symbiont genomes**. *ISME J* 2014, **8**:2490–2502.

77. Raina M, Ibba M: **tRNAs as regulators of biological processes**. *RNA* 2014, **5**:171.

78. Pinheiro PV, Kliot A, Ghanim M, Cilia M: **Is there a role for symbiotic bacteria in plant virus transmission by insects?** *Curr Opin Insect Sci* 2015, **8**:69–78. [*Ecology * Parasites/Parasitoids/Biological Control*]

FIGURE LEGENDS

Figure 1. Length distribution of sRNA reads from aphid samples aligned to *Myzus persicae Densovirus*.

Figure 2. Quantification of *Myzus persicae Densovirus* (MpDNV) in *Myzus*

persicae aphids (10 aphids per replicate/ 4 replicates per treatment) fed to potato plants cv. Red Maria that were infected with PLRV, mock inoculated or noninoculated (healthy) by quantitative PCR. Different letters show significantly different treatments ($p<0.05$).

Figure 3. Fluorescence *In Situ* Hybridization (FISH) of *Myzus persicae Densovirus* (MpDNV) and *Potato leafroll virus* (PLRV) in guts of PLRV-viruliferous (a, b, e and f) and PLRV-free (c, d, g and h) *Myzus persicae* aphids. Blue in all panels is DAPI staining of the nuclei. Red is staining of PLRV using FISH probe specific sequence conjugated to Cy2. Green is staining of MpDNV using specific FISH probe conjugated to Cy3. Fg: foregut; mg: midgut.

Figure 4. Distribution of siRNA reads along the *Myzus persicae Densovirus* (MpDNV) genome in aphids fed to PLRV-infected potato, PLRV-containing artificial diet, mock-inoculated potato and PVY-infected potato. A) 22nt long siRNA reads; B) 34-38 nt long siRNA reads.

Figure 5. Size distribution of small interfering RNA (siRNA) generated in *Myzus persicae* that aligned to the *M. persicae* genome.

Figure 6. Number of small interfering RNA (siRNA) derived from *Potato leafroll virus* in potato samples infected with PLRV at three days post inoculation (3 DPI) and at three weeks post inoculation (3 WPI).

Figure 7. Small interfering RNA (siRNA) derived from *Buchnera aphidicola* Asparagine-tRNA and generated by *Myzus persicae*.

Figure 8. Small interfering RNA (siRNA) derived from *Buchnera aphidicola* aminoacyl-tRNAs and generated by *Myzus persicae*.

SUPPLEMENTAL MATERIAL

Figure S1. Distribution of sRNAs generated in aphids fed on A) PLRV-infected potato plants, B) purified PLRV in 30% sucrose, C) Mock-inoculated potato plants and D) 30% sucrose, along the *Myzus persicae Densovirus* (MpDNV) genome.

Table S1. Alignment of *Myzus persicae* (clone G006 aphidbase.org) and *Buchnera aphidicola* (F009 strain from *Myzus persicae*, CP002703) genomes resulted in eight regions with high similarity, based on a e-value cut-off of $> 10^{-5}$.

Table S2. *Buchnera* tRNA-derived small RNA in aphid and plant samples by tRNA gene, total number of reads, average of reads across replicates and percentage of total *Buchnera* tRNA-derived sRNAs. (Spreadsheet file, to be added as supplemental data)

Table 1. Summary of Illumina deep sequencing data.

Tissue	Treatment	Total # reads	# Reads after removing adapters and rRNA	PLRV	PVY	MpDNV	*Buchnera aphidicola*	*Buchnera* tRNAs (% of total *Buchnera* reads)
Aphids	Aphids fed on PLRV-infected potato + turnip	31929115	5673826	64.5	79.6	2979.7	471394.7	51058.3 (10.8%)
Aphids	Aphids fed on purified virus + turnip	29967834	5966972	48.5	108.3	2845.7	723696	99150 (13.1%)
Aphids	Aphids fed to mock inoculated potato + turnip	7370826	1898702	--	51.3	2196	633184	26214 (4.1%)
Aphids	Aphids fed to PVY-potato + turnip	13335948	4909636	--	203	6830	1594164	117013 (7.3%)

Aphids	Aphids fed to sucrose only diet	5083965	1735400	--	14	6457	447774	66272 (14.8%)
Aphids	Aphids fed to sucrose only diet + turnip	9036073	3270474	--	48.7	13340	1182707	147531 (12.5%)
Potato leaves	PLRV-infected potato 3 weeks after inoculation	8231543.7	1885542	21918.3	13	79.6	9187	271 (2.9%)
Potato leaves	PLRV-infected potato 3 days after inoculation	3521369	1317272	3419	3	--	2203	186 (8.4%)
Turnip leaves	Turnip fed to aphids after feeding on PLRV-infected potato	8426652	1158558	108	33.3	37	12988.3	945 (7.2%)
Turnip leaves	Turnip fed to aphids after feeding on purified PLRV	14625639.7	1873005	169	42.3	48	9283.7	397 (4.2%)

Figure 1.

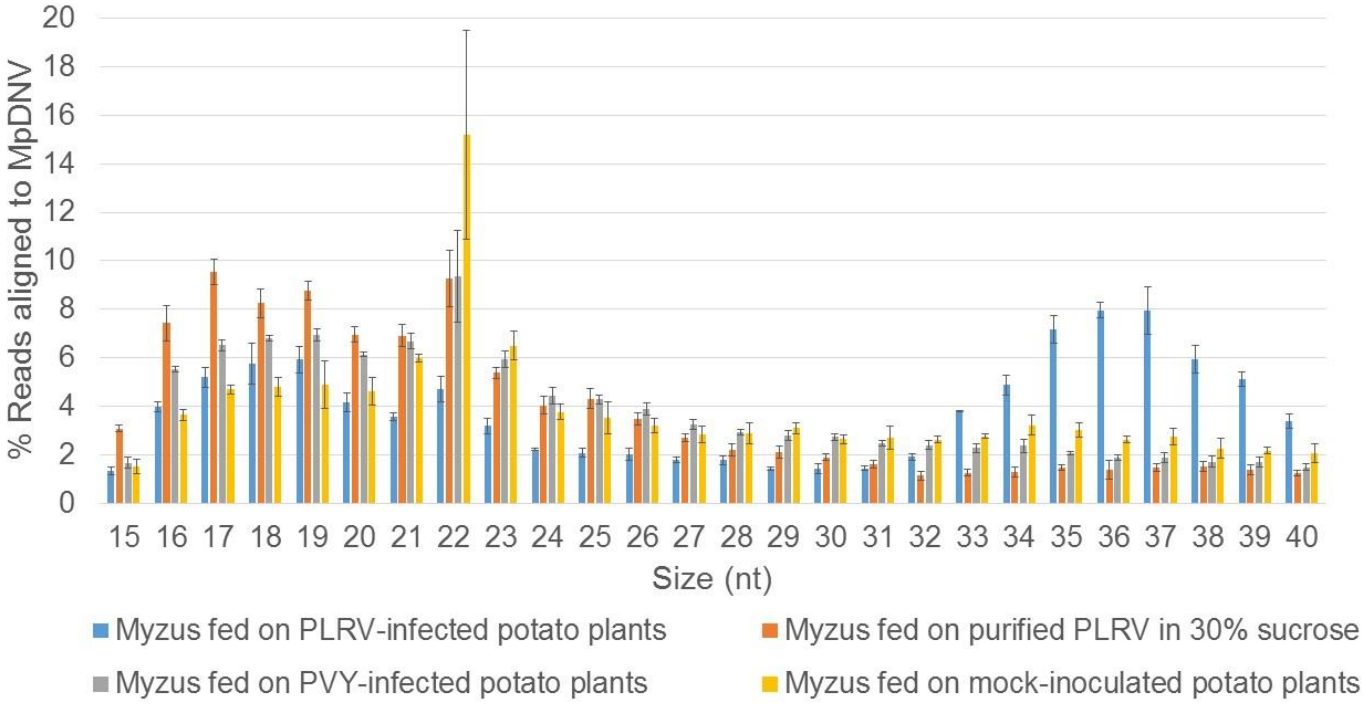

% Reads aligned to MpDNV
20
18
16
14
12
10
8
6
4
2
0
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Size (nt)
Myzus fed on PLRV-infected potato plants
Myzus fed on purified PLRV in 30% sucrose
Myzus fed on PVY-infected potato plants
Myzus fed on mock-inoculated potato plants

Figure 2.

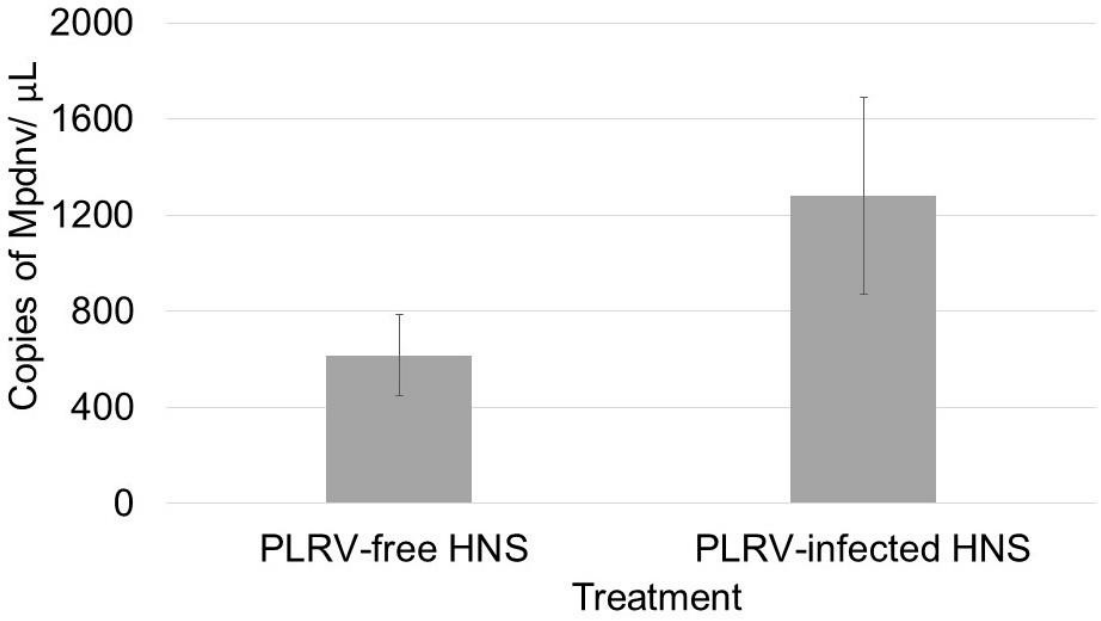

2000
1600
1200
800
400
0
Copies of Mpdnv/ µL
PLRV-free HNS
PLRV-infected HNS
Treatment

Figure 3.

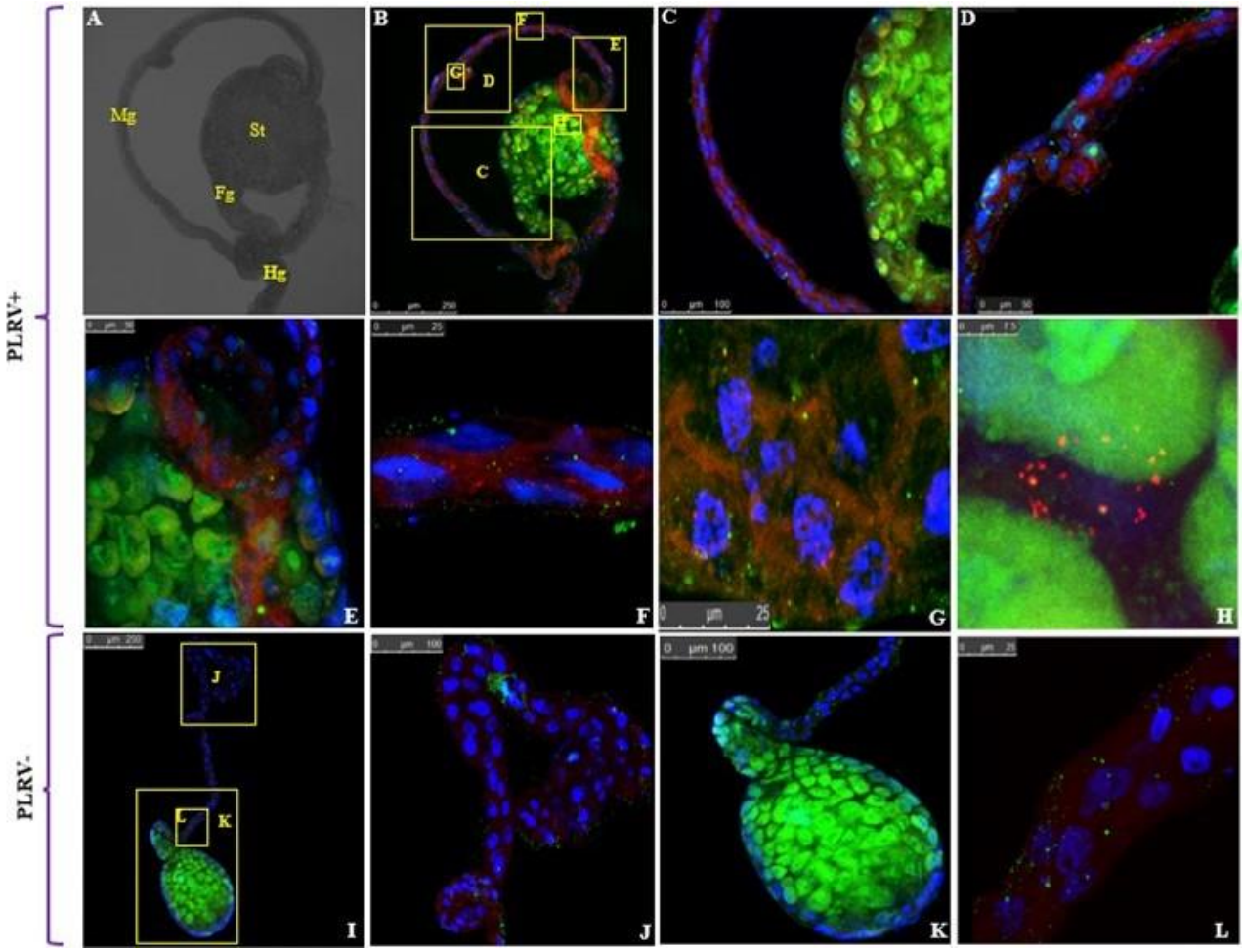


Figure 4

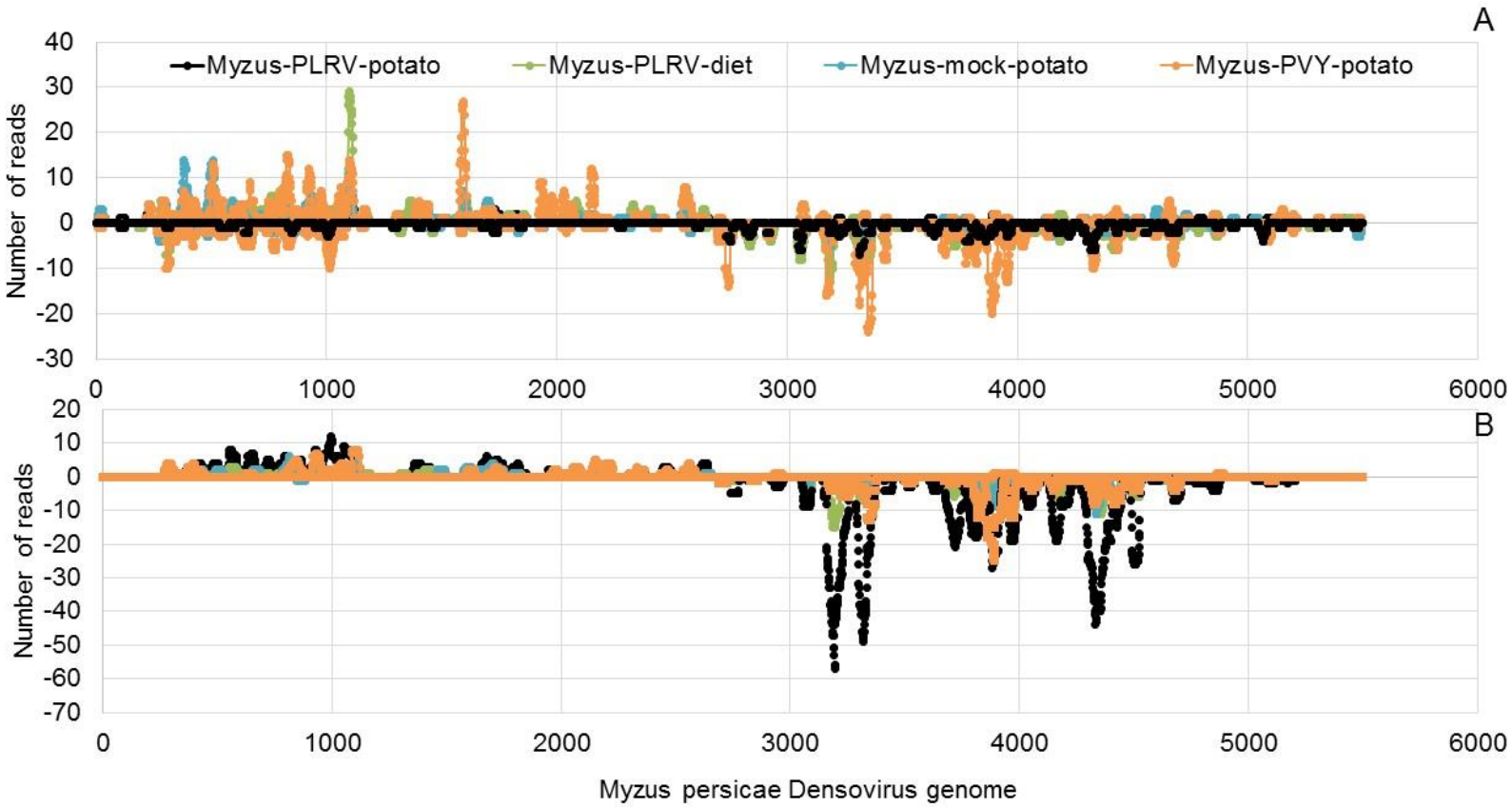

A
Myzus-PLRV-potato
Myzus-PLRV-diet
Myzus-mock-potato
Myzus-PVY-potato
Number of reads
40
30
20
10
0
-10
-20
-30
0
1000
2000
3000
4000
5000
6000
B
Number of reads
20
10
0
-10
-20
-30
-40
-50
-60
-70
0
1000
2000
3000
4000
5000
6000
Myzus persicae Densovirus genome

Figure 5

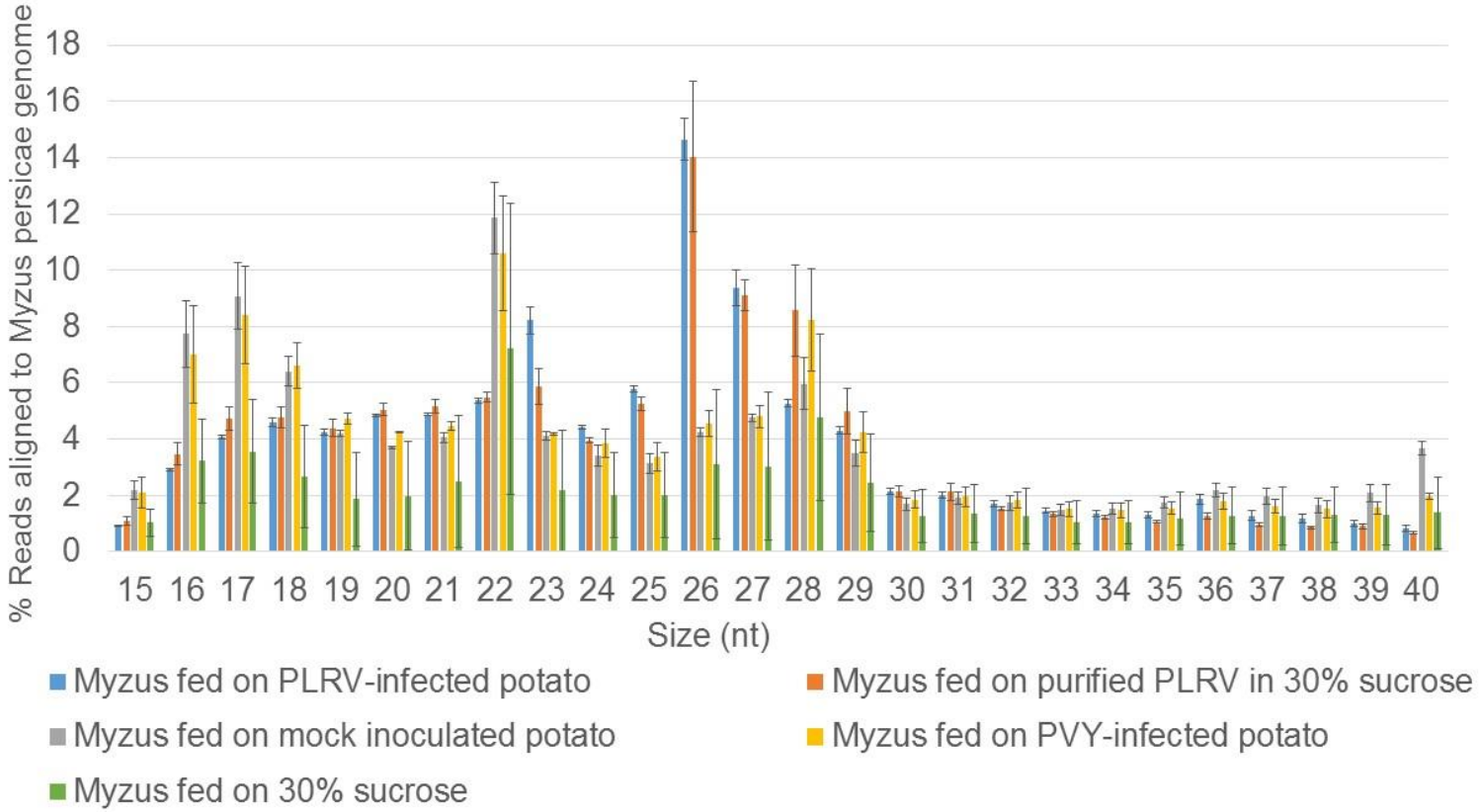

% Reads aligned to Myzus persicae genome
0
2
4
6
8
10
12
14
16
18
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Size (nt)
Myzus fed on PLRV-infected potato
Myzus fed on purified PLRV in 30% sucrose
Myzus fed on mock inoculated potato
Myzus fed on PVY-infected potato
Myzus fed on 30% sucrose

Figure 6. Number of sRNA reads from potato samples aligned to the *Potato leafroll virus* genome at three days post inoculation (3 DPI) and at three weeks post inoculation (3 WPI).

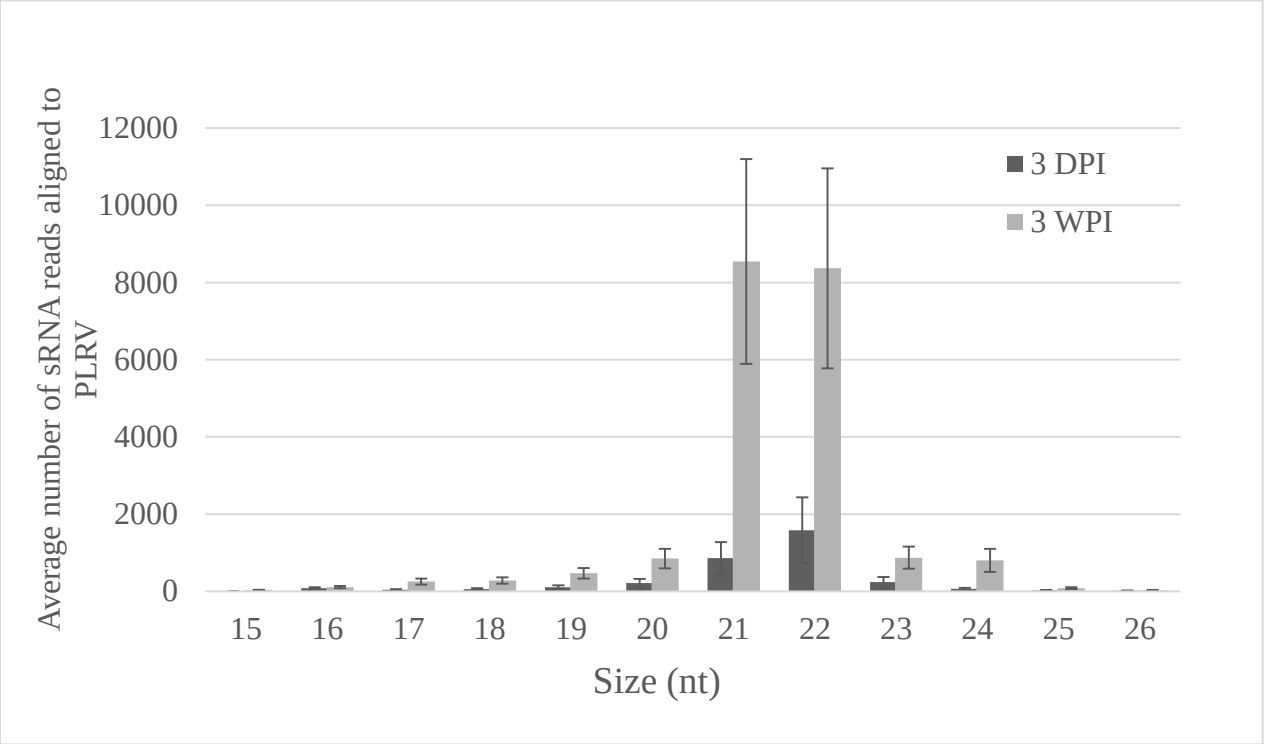


Figure 7. Aphid sRNA reads aligned to the *Buchnera aphidicola* Asparagine-tRNA.

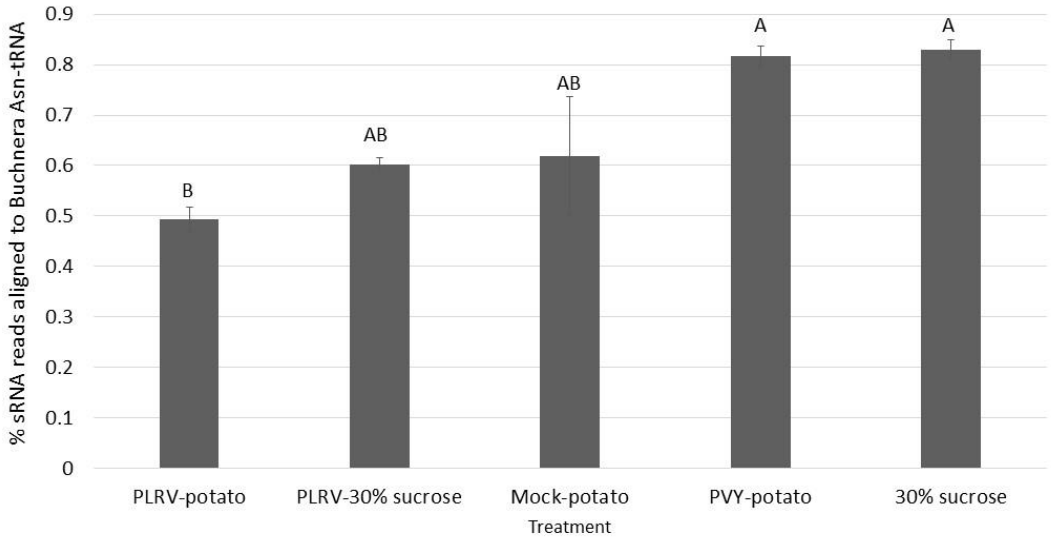
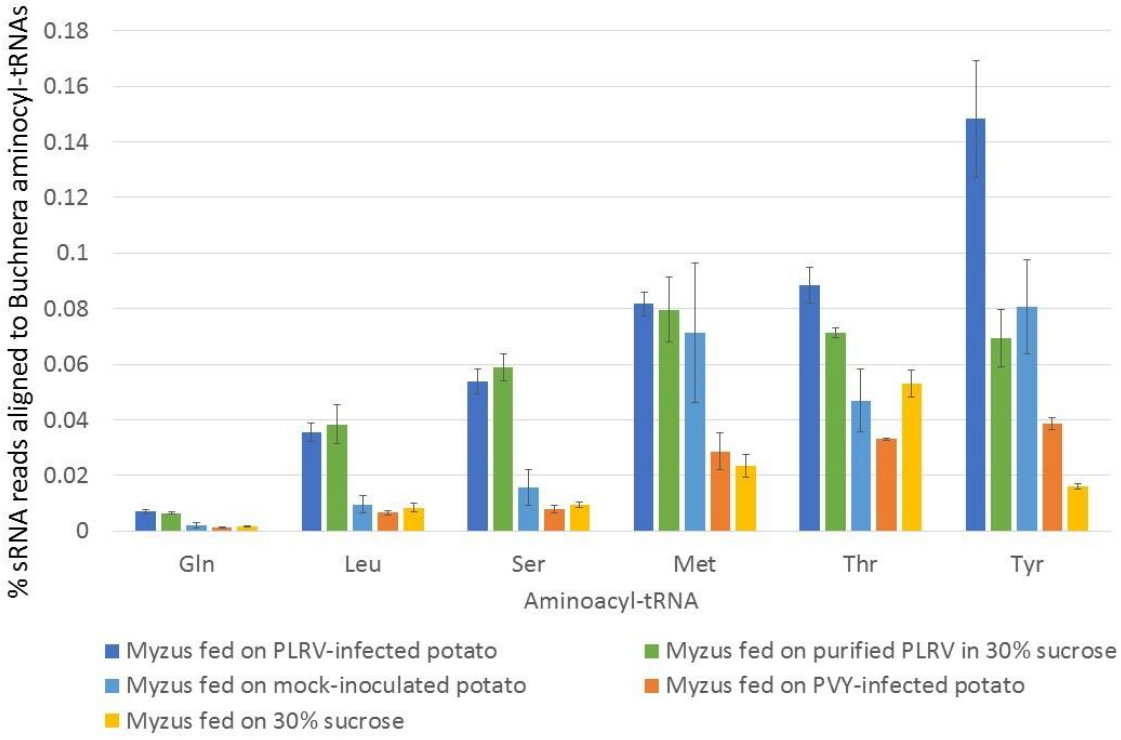


Figure 8. Aphid sRNA reads aligned to *Buchnera aphidicola* aminoacyl-tRNAs.



Appendices

I. The conditional vectoring hypothesis

II. The draft genome of the whitefly Bemisia tabaci (*as a contributing author*)

Appendix I

The conditional vectoring hypothesis

Unpublished work

Our study presented in chapter 2 shows that there is a host effect on the regulation of gut cysteine proteases in *M. persicae* that directly or indirectly impairs the transmission of PLRV by the aphid. The question that remains is why? Aphids can be broadly classified into two groups: as specialists with near exclusive niches, or generalist aphids with polyphagous feeding habits. The variation in luteovirid transmission efficiency in specialist aphids collected from the field, for example *Schizaphis graminum*, is massive, ranging from zero to 100% efficient, and is primarily due to a few aphid genes and protein isoforms that are additive in effect (53, 55, 88–90). Although several proteins have been identified to interact with plant viruses in generalist insect vectors (reviewed in (3)), the genetic basis for vectoring capacity in generalist aphids, such as *M. persicae*, and in other generalist insect plant virus vector species, such as the whitefly *Bemisia tabaci*, is less clear. Studies have documented positive fitness effects of virus-infected plants on aphids, including *M. persicae* and the potato aphid *Macrosiphum euphorbiae* feeding on PLRV-infected plants (91). We present a new hypothesis we call the conditional vector strategy hypothesis, not presented in the published article, where the virus transmission efficiency of a generalist insect vector of plant viruses is modulated according to interactions with

host plants. Such a conditional vectoring strategy may save the aphid the cost of harboring a plant virus in its body when the recipient host plant is not also a host of the virus. The hypothesis predicts that during interactions where the host plant is a host of both the insect and the virus, the insect vector will modulate the expression of genes and proteins to increase virus transmission efficiency and modulate reproductive output when it benefits the insect vector the most. When the aphid's host plant is not a host of the virus, the aphid receives no benefit from transmitting the virus and will reduce its investment in transmission. The conditional vectoring strategy idea can be tested empirically in the future and presents new strategies for controlling transmission in the field.

Appendix II

The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance

Published as a research article in BMC Biology (2016) 14: 110, DOI: 10.1186/s12915-016-0321-y

Wenbo Chen[1,†], Daniel K. Hasegawa[1,2,†], Navneet Kaur[3], Adi Kliot[4], Patricia Valle Pinheiro[1,5,6], Junbo Luan[6], Marcus C. Stensmyr[7], Yi Zheng[1], Wenli Liu[1], Honghe Sun[1], Yimin Xu[1], Yuan Luo[6], Angela Kruse[1,8], Xiaowei Yang[6], Svetlana Kontsedalov[4], Galina Lebedev[4], Tonja W. Fisher[9], David R. Nelson[10], Wayne B. Hunter[11], Judith K. Brown[8], Georg Jander[1], Michelle Cilia[1,7,12], Angela E. Douglas[6], Murad Ghanim[4], Alvin M. Simmons[2], William M. Wintermantel[3,*], Kai-Shu Ling[2,*], Zhangjun Fei[1,12,*]

[1]Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
[2]U.S. Department of Agriculture-Agricultural Research Service, U.S. Vegetable Laboratory, Charleston, SC 29414, USA
[3]U.S. Department of Agriculture-Agricultural Research Service, Crop Improvement and Protection Research, Salinas, CA 93905, USA
[4]Department of Entomology, The Volcani Center, Bet Dagan 50250, Israel
[5]EMBRAPA Rice and Beans, Santo Antônio de Goiás, GO, 75375-000, Brazil
[6]Department of Entomology, Cornell University, Ithaca, NY 14853, USA
[7]Department of Biology, Lund University, SE-223 62 Lund, Sweden
[8]Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, USA
[9]Department of Plant Sciences, The University of Arizona, Tucson, AZ 85721, USA
[10]Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, TN 38163, USA
[11]U.S. Department of Agriculture-Agricultural Research Service, U.S. Horticultural Laboratory, Fort Pierce, FL 34945, USA
[12]U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

†Equal contributor.
*Correspondence: zf25@cornell.edu, kai.ling@ars.usda.gov, bill.wintermantel@ars.usda.gov

Abstract

Background: The whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) is among the 100 worst invasive species in the world. As one of the most important crop pests and virus vectors, *B. tabaci* causes substantial crop losses and poses a serious threat to global food security.

Results: We report the 615 Mb high-quality genome sequence of *B. tabaci* Middle East-Asia Minor 1 (MEAM1), the first genome sequence in the Aleyrodidae family, which contains 15,664 protein-coding genes. The *B. tabaci* genome is highly divergent from other sequenced hemipteran genomes, sharing no detectable synteny. A number of known detoxification gene families including cytochrome P450s and UDP-glucuronosyltransferases are significantly expanded in *B. tabaci*. Other expanded gene families including cathepsins, large clusters of tandemly duplicated *B. tabaci*-specific genes, and phosphatidylethanolamine-binding proteins (PEBPs), were found to be associated with virus acquisition and transmission and/or insecticide resistance, likely contributing to the global invasiveness and efficient virus vectoring capacity of *B. tabaci*. The presence of 142 horizontally transferred genes from bacteria or fungi in the *B. tabaci* genome, including ones encoding hopanoid/sterol synthesis and xenobiotic detoxification enzymes that are not present in other insects, offers novel insights into the unique biological adaptations of this insect such as polyphagy and insecticide resistance. Interestingly, two adjacent bacterial pantothenate biosynthesis genes, *panB* and *panC*, have been co-transferred into *B. tabaci* and fused into a single gene that has acquired introns during its evolution.

Conclusions: The *B. tabaci* genome contains numerous genetic novelties, including expansions in gene families associated with insecticide resistance, detoxification and virus transmission, as well as numerous horizontally transferred genes from bacteria and fungi. We believe these novelties likely have shaped *B. tabaci* as a highly invasive polyphagous crop pest and efficient vector of plant viruses. The genome serves as a reference for resolving the *B. tabaci* cryptic species complex, understanding fundamental biological novelties, and provides valuable genetic information to assist the development of novel strategies for controlling whiteflies and the viruses they transmit.

Keywords: whitefly, *Bemisia tabaci*, draft genome, virus transmission, polyphagy, insecticide resistance

Background

Whiteflies are notorious agricultural pests that have become major threats to global food security, and cause damage to crops by direct feeding and efficient transmission of numerous viruses infecting food, fiber and ornamental crops worldwide. Among the 1,556 known whitefly species in 161 genera [1], *Bemisia tabaci* (Hemiptera: Aleyrodidae) is particularly important because of its ability to infest more than 1,000 plant species [2] and transmit over 300 plant pathogenic viruses [3]. Major crops impacted by *B. tabaci*-transmitted viruses on a global scale include tomato, cassava, cotton, cucurbits, sweetpotato, and numerous other species. *Bemisia tabaci*-transmitted *Tomato yellow leaf curl virus* (TYLCV) causes one of the most devastating diseases affecting tomato production [4], and has spread globally [5], while outbreaks of cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) have reached epidemic levels in Africa [6-8] and are so severe that the global alliance on cassava virus research has declared a war against whiteflies and the viruses they transmit [9]. Furthermore, increasing global commodity trade, climate change, and intensive crop production are facilitating both the global dispersal and the development of super-abundant populations of *B. tabaci*, one of the 100 worst invasive alien species in the world (http://www.issg.org).

Bemisia tabaci was first identified as a new pest species in 1889 in Greece [10] and is now recognized to comprise multiple genetic groups, also known as “biotypes” [11]. Early work that assigned *B. tabaci* to various biotypes on the basis of several biological properties such as host range, behavior, insecticide resistance and virus transmission capacity [12, 13] has been replaced by more reliable molecular criteria for *B. tabaci* differentiation. For example, at least 34 genetic groups (or cryptic species) of *B. tabaci* have been discriminated based on the sequence divergence of the mitochondrial cytochrome oxidase I (MtCOI) gene [14-17], including two globally-important pest taxa, Middle East-Asia Minor 1 (MEAM1, formerly biotype B) and Mediterranean (MED, formerly biotype Q) [18].

Despite its agronomic importance, genomic resources for the *B. tabaci* whitefly are limited. Multiple transcriptome data are available, addressing the phylogenetic relationship and transcriptome sequence divergence of different *B. tabaci* species [19, 20], responses to a begomovirus [21], insecticide resistance [22], development and organ-specific patterns of gene expression [23-25], and the interactions with symbiotic bacteria required by the insect [26]. However, a fully sequenced *B. tabaci* genome is still greatly needed for further resolution of the species complex conundrum. In addition, a reference genome will assist our understanding of the molecular mechanisms underlying virus transmission, detoxification, host adaptation, and insecticide resistance.

Here, we present a high-quality draft genome sequence of *B. tabaci* MEAM1, which was assembled using a hybrid approach involving Illumina short reads and PacBio long reads. This assembly represents the first genome sequence of a member of the family Aleyrodidae. The availability of the *B. tabaci* genome not only provides novel insights into the underlying mechanisms of the whitefly’s global invasiveness and high virus vector capacity, but also presents valuable information to help understand the *B. tabaci* species complex and to facilitate the development of improved strategies for efficient whitefly management.

Results and Discussion

The genome of *B. tabaci*

Whiteflies from a *B. tabaci* colony established from a single female collected at the USDA-ARS in Charleston, SC were used for genome sequencing (**Fig. 1A and Additional file 1: Figure S1**). PCR analysis using primers against the mitochondrial cytochrome oxidase I (MtCOI) gene [27] indicates that the colony is a member of the MEAM1 species. A total of 203.8 Gb high-quality cleaned Illumina sequences and 4 Gb PacBio long reads were generated (**Additional file 2**), which represented ~300-fold coverage of the *B. tabaci* MEAM1 genome that has an estimated size of ~690 Mb [28]. *De novo* assembly using Illumina and PacBio sequences resulted in a final draft genome of 615.0 Mb with an N50 scaffold size of 3.23 Mb, which spanned 89.1% of the *B. tabaci* genome (**Table 1**). Quality evaluation using BUSCO [29] revealed that 96.8% of the core eukaryotic genes were captured by the *B. tabaci* genome assembly and 94.4% were complete. In addition, the high mapping rates of the published whitefly mRNA sequences as well as our paired-end RNA-Seq reads further supported the high quality of the *B. tabaci* genome assembly (**Additional files 3** and **4**).

A total of 276.9 Mb (45%) of repeat sequences were identified in the *B. tabaci* genome, which is slightly higher than that of the related hemipteran *Acyrthosiphon pisum* genome (38%) [30]. Among these repeats, ~170.5 Mb (28%) were annotated as MITEs, while 79.7 Mb (13%) could not be classified into any known families (**Additional file 5**). A total of 15,664 protein-coding genes were predicted in the *B. tabaci* genome, among which 13,562 (87%) were supported by our RNA-Seq data, 7,321 (47%) by homologous proteins, and 6,473 (41%) by both. Of these, 81% were functionally annotated (**Additional file 6**). Despite the different sizes of the assembled *B. tabaci* (615.0 Mb) and *D. melanogaster* (142.6 Mb) genomes, the number of protein-coding genes in the two species was similar (15,664 vs 13,920). The mean coding sequence length of the genomes was also similar, while the mean intron and untranslated region (UTR) lengths in *B. tabaci* were considerably larger than those in *D. melanogaster* (**Additional file 7**).

Genome-based phylogeny and genome comparisons

We compared *B. tabaci* protein-coding genes with those of five exopterygotan insects, eight endopterygotan insects, and two non-insect arthropod species (**Additional file 8**) to identify orthologous groups. The phylogeny of these 16 species, based on 642 single-copy orthologous genes, shows that *B. tabaci* is a sister taxon to *A. pisum* (pea aphid), forming a lineage together with three other hemipteran insects, *Nilaparvata lugens* (brown planthopper), *Rhodnius prolixus* (Triatomid bug) and *Diaphorina citri* (Asian citrus psyllid) (**Fig. 1B**). Interestingly, no syntenic blocks were identified between any of these

hemipteran genomes. This is different from the Lepidoptera *Heliconius melpomene* (butterfly), *Bombyx mori* (silkworm) and *Plutella xylostella* (diamondback moth), whose genomes share high synteny [31]. Our analysis suggests that genomes of the five hemipteran insects, *B. tabaci*, *A. pisum*, *N. lugens*, *R. prolixus* and *D. citri,* are highly divergent, consistent with previous reports suggesting that *B. tabaci* and *A. pisum* diverged about 250 million years ago [32] whereas *H. melpomene* and *B. mori* diverged ~103 million years ago [33].

Among the 15,664 genes in the *B. tabaci* genome, 10,334 (8,372 gene families) had detectable homologues in the other 15 arthropods, including 2,817 (2,427 gene families) that were conserved in all 16 species (**Fig. 1B**). A total of 5,330 genes (3,885 gene families) including 3,417 single copy genes were found to be unique in *B. tabaci*. Furthermore, a total of 18 protein domains, which represented 10 gene families, were found to be significantly expanded in *B. tabaci* (**Fig. 1C and Additional file 9**). These expanded gene families include ones that are potentially involved in virus transmission or insecticide resistance, in addition to those that were horizontally transferred (see discussions below).

Vector for plant virus transmission

Bemisia tabaci is one of the most prevalent and agriculturally important vectors of plant viruses, capable of transmitting viruses from at least five genera, in a persistent circulative, semipersistent, or nonpersistent manner [34]. We compared transcriptome profiles of whiteflies during the first three days of virus acquisition feeding on tomato plants infected with TYLCV (genus *Begomovirus*), which is transmitted by *B. tabaci* in a persistent circulative manner, or *Tomato chlorosis virus* (ToCV, genus *Crinivirus*), which is transmitted in a semipersistent, non-circulative manner, to the corresponding whiteflies feeding on virus-free tomato plants for the same time periods. We found that during the acquisition feeding of TYLCV or ToCV-infected tomato plants, a large number of cathepsin genes were differentially expressed including 20 cathepsin B, five cathepsin L-like, three cathepsin F and one cathepsin F-like genes (**Fig. 2A and Additional file 10**). Cathepsins are proteases involved in many biological processes, including protein degradation, apoptosis and signaling, and their activity in the late endosome and lysosome has been widely implicated in virus transmission [35, 36]. A total of 111 cathepsin genes were detected in the *B. tabaci* genome (**Fig. 2B**), representing a significant expansion when compared to the other 15 arthropod species that were examined **(Additional file 11)**. Specifically, a large expansion of cathepsin B genes was observed, with 50 members identified, many of which were tandem duplications. In addition, the *B. tabaci* genome contains 35 cathepsin L-like genes, while none were found in the genomes of the other 15 arthropods, indicating that these unique cathepsin L-like genes represent a novel *B. tabaci*-specific clade of cathepsins (**Fig. 2B** and **Additional file 11**). The expansion of cathepsin B and L-like families in *B. tabaci* could be tied to the tremendous efficiency of this insect species as a vector of numerous and diverse plant viruses, possibly through its involvement in immune responses to virus acquisition or other responses that govern whitefly-virus interactions.

Interestingly, three large clusters in the *B. tabaci* genome were found to contain *B. tabaci*-specific unknown genes that were differentially expressed during acquisition feeding of *B. tabaci* on ToCV-infected tomato plants. Most of these genes were tandem duplications (**Fig. 2C and Additional file 1: Figure S2**). Our results suggest that during the evolution of *B. tabaci*, these specific genomic regions might have contributed to the elevated ability of this whitefly to transmit plant viruses, particularly non-circulative, semipersistent viruses, since these genes were not differentially expressed during feeding on tomatoes infected with the persistent, circulative virus, TYLCV. The differential expression of these unique clusters in specific association with virus acquisition feeding on ToCV-infected tomato indicates a response by the whitefly to either ToCV itself or to host factors uniquely expressed in the tomato plant during infection by ToCV. Although no function has been attributed to these genes, their expression during acquisition of ToCV from infected tomato plants suggests they may represent genes that are co-evolved in the whitefly vector that facilitate uptake, retention, or transmission of ToCV and perhaps other semipersistent viruses.

Detoxification and insecticide resistance

Bemisia tabaci is highly polyphagous, being able to feed on more than 1,000 different plant species, and is notable for its rapid development of resistance to numerous insecticides. Thus, *B. tabaci* likely have developed the capacity to overcome a wide variety of plant defense compounds and insecticides. Several enzyme families implicated in detoxification were identified in the *B. tabaci* genome, including cytochrome P450s (CYPs), UDP-glucuronosyltransferases (UGTs), glutathione S-transferases (GSTs), ABC transporters (ABC) and carboxylesterases (CCEs) (**Additional file 12**). The *B. tabaci* genome contains 130 CYPs, representing a significant expansion relative to most insects with genomes sequenced. Notable expansions include a novel family (CYP3133) with 20 members, the CYP4CS subfamily with 14 genes and the CYP402C subfamily with 12 members (**Additional file 13** and **Additional file 1: Figure S3**). The *B. tabaci* genome encodes 81 UGTs, similar to *Tetranychus urticae* (81) and *A. pisum* (72), but substantially more than that are found in other insects (4 to 38). Additionally, twenty-two GSTs (**Additional file 1: Figure S4**), 50 ABCs (**Additional file 1: Figure S5**) and 51 CCEs genes were detected in the *B. tabaci* genome. Expansion of some of these detoxification gene families in *B. tabaci* likely provides a basis for its well-known insecticide resistance and its ability to occupy a broad range of host plants with a diversity of defenses.

Currently, the MEAM1 and MED cryptic species of *B. tabaci* are the most widely prevalent throughout the world, and have greatly expanded their ranges over the past two decades, with MED having developed broader insecticide resistance than MEAM1 [37]. We compared global transcriptome profiles of a susceptible MED population (PyriR), as well as a resistant MED population (9-2013), with and without treatment with the insecticide Mospilan (acetamiprid). As expected, all of the aforementioned detoxification families contained genes that were responsive to Mospilan treatment in both susceptible and resistant populations, supporting their roles in whitefly insecticide resistance (**Fig. 3A** and **Additional file 14**). Interestingly, a large number of genes from the highly expanded

cathepsin family were differentially expressed upon Mospilan treatment, with 26 and 12 in susceptible and resistant populations, respectively. Cathepsins have been associated with the polyphagous habit of the whitefly [38]. This and the novel role of cathepsins in insecticide resistance revealed here, suggest that cathepsins might have contributed to the global invasiveness of the whitefly.

In addition, the phosphatidylethanolamine-binding protein (PEBP) gene family, which has not been previously associated with detoxification or insecticide resistance in insects, showed striking responses to Mospilan treatment. A total of 134 and 16 PEBP genes were responsive to Mospilan treatment in the MED resistant and susceptible populations, respectively, all of which were down-regulated (**Fig. 3A** and **Additional file 14)**. PEBPs are a highly conserved group of proteins that have been identified in a wide variety of organisms [39] and associated with various biological processes, including neuronal development [40], serine protease inhibition [39], and regulation of MAP kinase [41] and NF-kappaB [42] signaling pathways. Our analysis supports a novel and very important role for the PEBPs in insect resistance to pesticides. The *B. tabaci* genome contained 202 PEBPs, representing a tremendously expanded gene family and containing several new clades/subfamilies (**Fig. 3B**). By comparison, the genomes of the other 15 arthropods had a maximum of 16 PEBPs. Among the *B. tabaci* PEBPs, 127 were located in five large tandem clusters, the majority of which were responsive to insecticide treatment (**Fig. 3C** and **Additional file 1: Figure S6**). Our data suggest a strong role for PEBPs in *B. tabaci* insecticide resistance and that the large expansion of this family may have contributed to its rapidly evolved insecticide resistance.

Endosymbiont genomes

Whiteflies harbor endosymbiotic bacteria, comprising a primary symbiont *Portiera aleyrodidarum* and one or more additional bacteria, generically known as secondary symbionts [43]. Diagnostic PCR assays using the primers described in Pan et al. [43] indicated that the colony of MEAM1 used for genome sequencing bore the primary endosymbiont, *Portiera*, and two secondary endosymbionts, *Hamiltonella* and *Rickettsia*. The genomes of the three endosymbionts were assembled *de novo*, with each assembled into a single contig. The assembled genome of *Portiera* was complete with a size of 352 kb, while those of *Hamiltonella* and *Rickettsia* were nearly complete, with sizes of 1.74 Mb and 1.38 Mb, respectively (**Additional file 15** and **Additional file 1: Figure S7**). In *Portiera*, 273 genes were predicted, suggesting that it has a highly reduced genome largely comprising genes essential for basic cellular processes and whitefly nutrition. In contrast, 1,627 and 1,347 genes were predicted in *Hamiltonella* and *Rickettsia*, respectively. *Hamiltonella* possesses 94 (5.8%) phage genes and a large number of genes involved in the type II/III secretion systems. Approximately 22% of the *Rickettsia* genes are homologous to transposable elements, suggesting that the genome is highly dynamic. Comparative analysis of the *B. tabaci* genome with the *Portiera* and *Hamiltonella* genomes identified genes coding for complementary reactions in multiple metabolic pathways, including essential amino acid biosynthesis (**Additional file 16** and **Additional file 1: Figure S8**), as reported previously [24, 44]. Analysis of the *Rickettsia*

genome also shows the absence of genes for non-essential amino acid biosynthesis (**Additional file 16** and **Additional file 1: Figure S8**). Neither *B. tabaci* nor any of the endosymbiont bacteria appear to encode known enzymes that catalyze the conversion of histidinol to histidine, suggesting that one or more of these organisms might contain a non-canonical enzyme for the final step of histidine biosynthesis. The biosynthetic pathway leading from homoserine to methionine is incomplete in *B. tabaci* and its endosymbionts. However, *B. tabaci* does encode homocysteine methyltransferase, an enzyme that produces methionine from S-methylmethionine, one of the most abundant sulfur transport molecules in plants [45]. The homocysteine necessary for this reaction can be produced as a by-product of the S-adenosylmethionine cycle, which is present in *B. tabaci* and its endosymbionts. Almost all genes of the branched-chain amino acid biosynthesis pathways are present in *Portiera*. It is notable that branched chain amino acid aminotransferase, the only gene missing in *Portiera*, is present in both *B. tabaci* and *Rickettsia*, indicating that these two organisms can independently produce leucine, isoleucine, and valine from the respective oxo-acids.

Genes acquired horizontally from bacteria and fungi

The recent rapid accumulation of genomic data has facilitated the identification of increasing numbers of horizontally acquired exogenous DNA sequences in the genomes of animals, including insects [46]. We identified 142 horizontal gene transfers (HGTs) in the *B. tabaci* genome, with 64 of bacterial origin (**Additional file 17**) and 78 of fungal origin (**Additional file 18**). Recent reports on HGTs in the tardigrade genomes [47, 48] have demonstrated the importance of carefully examining eukaryotic genome assemblies to distinguish contaminants from authentic HGTs. In this study, we provide multiple lines of evidence to support the identified HGTs, including the alignments of paired-end and mate-pair DNA reads and polyA enriched strand-specific RNA-Seq reads (see Methods for details; **Additional files 17 and 18; Additional file 1: Supplementary Figure S9)**. In addition, our RNA-Seq data indicated that most of the HGTs were moderately or highly expressed, and 10 HGTs of bacterial origin were previously confirmed by qPCR [24]. Together, our data strongly support the high-confidence of the identified HGTs in the *B. tabaci* genome.

The majority of the *B. tabaci* HGTs (93) had predicted enzymatic functions. HGTs of bacterial origin mainly contributed to amino acid synthesis, vitamin synthesis, and lipid metabolism, while those of fungal origin mainly contributed to carbohydrate processes, pro-oxidant functions, and lipid metabolism. Two cases of co-transfer of two genes were identified: *bioA*-*bioD* phylogenetically allied with the bacterium *Cardinium* and *panB*-*panC* allied with the bacterium *Pseudomonas*, which encode enzymes in the biosynthesis pathways of biotin (vitamin B7) and pantothenate (vitamin B5), respectively. The *bioA* and *bioD* genes are adjacent to one another in the *Cardinium* genome; while in the *B. tabaci* genome they are arranged as two sets of adjacent genes, with the *bioA* truncated in one pair (Bta00841), and *bioD* truncated in the second pair (Bta01938) (**Additional file 1: Supplementary Figure S10**), suggesting that the genes were duplicated and pseudogenized due to functional redundancy. In the other case, *panB* and *panC* are two

adjacent genes in *Pseudomonas*, but become a single gene in the *B. tabaci* genome and have acquired introns (**Fig. 4A and Additional file 1: Supplementary Figure S9**). It has been reported that genes of bacterial origin can acquire introns after their transfer into eukaryotic genomes [49, 50], and a large portion of *B. tabaci* HGTs of bacterial origin also contain introns (**Additional files 17**). However, as far as we know, no reports have described that two adjacent bacterial genes might have been fused into one gene and acquired introns after horizontal transfer. This arrangement of *panB* and *panC* in *B. tabaci* likely promotes coordinated enzymatic functions. The PanB and PanC domains of the fused protein are predicted to mediate the proximal and final reactions in pantothenate synthesis. Neither *B. tabaci* nor its primary endosymbiont *Portiera* apparently possesses the canonical gene, *panE*, mediating the intermediate step. However, *Portiera* does have *ilvC*, which has been shown to mediate the *panE* reaction in another symbiotic bacterium (*Buchnera* in aphids) [51], suggesting that *B. tabaci-Portiera* association may be capable of pantothenate synthesis by a shared metabolic pathway between the horizontally-acquired gene in the insect genome and the symbiont gene (**Fig. 4B**).

The *B. tabaci* genome contains a gene of fungal origin annotated as squalene synthase (**Additional file 18**), which mediates the first committed reaction in sterol synthesis, and seven genes of bacterial origin coding for squalene hopene cyclases (**Fig. 4C** and **Additional file 17**), which are predicted to synthesize hopanoids, the bacterial analogs of sterols. In animals, including insects, sterols function to maintain the structural integrity of membranes and also act as hormones (e.g. the ecdysteroid molting hormones of insects) [52]. Although most animals can synthesize sterols, insects and other arthropods lack this metabolic capability and are generally dependent on a dietary supply of sterols [53]. The potential capacity of *B. tabaci* to synthesize sterols/hopanoids, which would negate their dietary requirement, may be of selective advantage given phloem sap has low sterol content [54], and may contribute to the exceptionally wide host range of this whitefly species.

We detected 20 aromatic peroxygenase (APO) genes of fungal origin in the *B. tabaci* genome, but none were present in any other insect genomes. APOs function in detoxification by selectively hydroxylating the aromatic ring of toxic compounds such as naphthalene [55]. In fungi, these enzymes have been implicated in the degradation of complex plant biomolecules [56]. One of the APOs was differentially expressed upon treatment with the insecticide Mospilan (**Fig. 3A**). We hypothesize that expression of the horizontally transferred APOs could contribute to the insecticide resistance of whiteflies as well as their high capacity for xenobiotic detoxification.

Conclusions

The whitefly *B. tabaci* represents one of the most agronomically significant pests. Our analysis of the *B. tabaci* genome also included chemosensory genes, immunity-related genes, and genes in the RNA interference (RNAi) pathway (**Additional file 1: Supplementary text**). Analyses of the *B. tabaci* genome reveal numerous genetic novelties that likely have shaped whiteflies as a highly invasive pest of agricultural crops

and as one of the most prevalent and efficient vectors of plant viruses. These include (1) several *B. tabaci*-specific gene clusters that are tandemly duplicated and uniquely responsive to feeding on virus-infected plants; (2) largely expanded gene families including cathepsins, CYPs, UGTs, and PEBPs that have potential roles in virus transmission, polyphagy, detoxification, and/or insecticide resistance; (3) a large number of genes horizontally transferred from bacteria and fungi, including those involved in essential amino acid and hopanoid/sterol synthesis, in addition to APOs with putative roles in detoxification. The *B. tabaci* genome reported here provides an important advance for understanding whitefly biology, with implications for insect pest management and associated virus control. Furthermore, the *B. tabaci* genome represents the first genome sequence in the Aleyrodidae family and is highly divergent from other sequenced hemipteran genomes, providing a valuable resource for future comparative and evolutionary genomic studies.

Methods

Genome sequencing, assembly and annotation

Genomic DNA was isolated from approximately 6,500 haploid male individuals from a *B. tabaci* MEAM1 colony established from a single female collected at the USDA-ARS in Charleston, South Carolina, USA in April, 2013, as described in Chen et al. [28]. The colony was validated as MEAM1 using primers specific to the mitochondrial cytochrome oxidase I (MtCOI) gene. Primer sequences used were: Btab-B (MEAM1) F:5'-CTAGGGTTTATTGTTTGAGGTCATCATATATTC-3', R:5'-AATATCGACGAGGCATTCCCCCT-3'; Btab-Q (MED) F:5'-CTTGGTAACTCTTCTGTAGATGTGTGTT-3', R:5'-CCTTCCCGCAGAAGAAATTTTGTTC-3'; Btab-NW (New World) F:5'-TACTGTTGRAATAGATGTTGACACTCGGG-3', R:5'-GGAAAAAATGTCAGRTTTACTCCCWCAAATATT-3', Btab-Uni (universal *Bemisia tabaci*) F:5'- GAGGCTGRAAAATTARAAGTATTTGG-3', R:5'-CTTAAATTTACTGCACTTTCTGCCAYATTAG-3' which amplified 478 bp, 303 bp, 405 bp, and 745 bp of the MtCOI gene, respectively [27]. PCR amplifications were performed in 20 ul reactions using GoTaq Green Master Mix (Promega, USA), 0.25 uM of each forward and reverse primer, and 150 ng DNA with initial denaturation at 95 ^{o}C for 2 m, 35 cycles of denaturation at 95 ^{o}C for 30 s, annealing at 46 ^{o}C (Btab-Uni) or 64 ^{o}C (Btab-B, -Q, -NW) for 1 m, extension at 72 ^{o}C for 1 m, and a final extension at 72 ^{o}C for 5 m. PCR products were visualized on a 1% agarose gel.

Three Illumina paired-end libraries with insert sizes of approximately 300 bp, 500 bp and 1 kb, and three Illumina mate-pair libraries, with insert sizes of 3-5 kb, 8-10 kb and 15-20 kb were constructed using the Genomic DNA Sample Prep kit and the Nextera Mate Pair Sample Preparation kit, respectively, following the manufacturer's instructions (Illumina, San Diego, CA). These libraries were sequenced on the Illumina HiSeq 2500

system. In addition, one PacBio library was prepared and sequenced on a total of 27 SMRT cells of Pacific Biosciences RSII Sequencing System using the P5C3 chemistry.

The Illumina reads were first processed to collapse duplicated read pairs into unique read pairs. Duplicated read pairs were defined as those having identical bases in the first 100 bp of both left and right reads. Illumina adaptor and low quality bases were trimmed from the reads using Trimmomatic [57]. Reads less than 40 bp were discarded. Errors in the Illumina sequencing reads were further corrected using Quake [58]. Sequencing errors in PacBio reads were corrected with PBcR [59] using the Illumina paired-end reads. For *de novo* assembly, the high-quality cleaned reads from the Illumina paired-end and mate-pair libraries were first assembled using Platanus [60] with parameters of "-s 5 -c 5 -u 0.2". Gaps within each scaffold in the resulting genome assembly were filled with Illumina paired-end reads using Gapcloser [61]. The error-corrected PacBio long reads were subsequently used to further fill gaps in the scaffolds and to connect scaffolds using PBJelly [62]. The assembled scaffolds were polished with iCORN2 [63] using paired-end Illumina reads to correct base errors. The assembled scaffolds were then aligned against the NCBI non-redundant nucleotide (nt) database using BLASTN with an e-value cut-off of 1e-5. Scaffolds with over 90% of their length similar to bacterial sequences were considered contaminants and removed. To remove further redundant sequences in the assembly, scaffolds were blasted against themselves and those contained within other scaffolds with sequence identity > 99% and coverage > 99% were removed.

Transcriptome sequencing and analysis

Eggs, nymphs, and pupae were collected from leaves of collard plants (*Brassica oleracea* L.) on which the isogenic MEAM1 colony was reared. Tissues were surface sterilized by submersion in a petri dish containing 70% ethanol. The eggs were gently separated from nymphs and pupae using a small paintbrush. Isolated nymph and pupa samples were rinsed with sterile water. Approximately 1,500 adult whiteflies reared on broccoli (*B. oleracea* L. var. *botrytis*) at the USDA-ARS in Charleston, SC were transferred to either TYLCV-infected or uninfected tomato (*Solanum lycopersicum* cv. Moneymaker) cuttings and allowed to feed for 24, 48, or 72 hours, respectively. For each treatment and time point, two compound leaves were collected from TYLCV-infected or uninfected plants and transferred to a flask filled with water, which was then sealed with Parafilm and placed in an insect proof cage. Whiteflies were added to each cage and allowed to feed for 24, 48, or 72 h under controlled conditions at 28 ± 1°C, 14:10 (L:D) h photoperiod, and ~60% humidity. A total of 200-500 living whiteflies were collected at the end of each time point and stored at -80°C until processing. Three biological replicates were performed for each sample. The same experiment under the same environmental conditions was performed using adults from a MEAM1 colony maintained at the USDA-ARS in Salinas, CA but were fed on ToCV-infected or uninfected tomato (cv. Moneymaker) plants.

For insecticide treatment experiments, adults of two MED populations, PyriR, which is susceptible to the insecticide Mospilan (acetamiprid), and 9-2103, which is

resistant, were fed on cotton seedlings (*Gossypium hirsutum* L. cv. Acala) treated with the insecticide Mospilan at an LC_{30} dose (lethal concentration required to kill 30% of the population; 2 ppm for PyriR and 100 ppm for 9-2013) with the dipping method, as previously described [64]. Whiteflies fed on untreated cotton seedlings were used as controls. The experiments were conducted under standard rearing room conditions of 25 °C, 50% relative humidity, and a light regime of 10 hours light and 14 hours dark. Three to four biological replicates, each containing a pool of 200-500 adult whiteflies were, collected from each treatment. The insects were kept at -80°C until use.

Total RNA was purified using the TRIzol Reagent (Ambion, USA) according to the manufacturer's protocol. Strand-specific RNA-Seq libraries were constructed following the protocol described in Zhong et al [65] and sequenced on the Illumina HiSeq 2500 system. Raw RNA-Seq reads were first processed to remove adaptor and low quality sequences using Trimmomatic [57]. Reads shorter than 40 bp after trimming were discarded. The resulting reads were then aligned to the ribosomal RNA database [66] and the three bacterial symbiont genomes using Bowtie [67], allowing up to three mismatches. The aligned reads were not used for further analysis. To assist gene prediction, the high-quality cleaned RNA-Seq reads were aligned to the assembled *B. tabaci* genome using Tophat [68] and the aligned reads were assembled into transcripts using Cufflinks [69]. For gene expression analysis, the RNA-Seq reads were aligned to the assembled *B. tabaci* genome using HISAT [70]. Raw counts for each *B. tabaci* predicted gene were derived from the read alignments and normalized to fragments per kilobase of exon model per million mapped fragments (FPKM). Differential expression analyses were performed using edgeR [71]. The resulting raw P values were adjusted for multiple testing using the false discovery rate (FDR) [72]. For each comparison, genes with FDR < 0.05 and fold change no less than 1.5 were considered as differentially expressed genes.

Annotation of repeat sequences

Repeat elements in the *B. tabaci* genome were first identified *de novo* using RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html), which integrates the output of RECON [73] and RepeatScout [74] to build, refine and classify consensus models of putative interspersed repeats. The resulting repeat sequences were aligned to the NCBI non-redundant protein (nr) database and those that were highly homologous to known proteins were removed. To identify repeat sequences in the *B. tabaci* genome, a library consisting of the *de novo* repeat elements identified by RepeatModeler and the Repbase library (http://www.girinst.org/repbase/index.html) were used to screen the assembled *B. tabaci* genome using RepeatMasker and RepeatRunner, which are integrated into the MAKER annotation pipeline [75]. Miniature inverted-repeat transposable elements (MITEs) were identified using MITE-Hunter [76].

Protein-coding gene prediction and annotation

Protein-coding genes in the *B. tabaci* genome were predicted with MAKER [75], which integrates the results from three different approaches: *ab initio*, homologous protein mapping and transcript mapping. Augustus [77] and SNAP [78] were used for *ab initio* gene prediction. For homologous protein mapping, protein sequences from the SwissProt database and the *Drosophila melanogaster* and *A. pisum* proteomes were aligned to the *B. tabaci* genome using Spaln [79] with default parameters. For transcript mapping, the *B. tabaci* mRNA sequences collected from GenBank were aligned to the genome using Spaln [79], and only mRNAs aligned to the genome with coverage greater than 90% and sequence identity greater than 97% were retained. In addition, the alignments of the reference-guided assembled transcripts from our RNA-Seq data, i.e. the GFF file generated by Cufflinks, were directly used by MAKER. From the *ab initio* predicted genes, MAKER generated a set of high confidence gene models, which were supported by transcript mapping and/or homologous protein mapping. The remaining *ab initio* predicted genes without evidence support were compared to the InterPro domain database [80] using InterProScan [81] and those containing InterPro domains were added into the predicted gene models. Finally, predicted gene models that overlapped with repeat sequences by 70% of their lengths were removed from the final predicted gene dataset.

The *B. tabaci* predicted genes were annotated by comparing their protein sequences against UniProt (TrEMBL and SwissProt), fruit fly and pea aphid proteomes, as well as the InterPro domain database. GO annotation was performed using Blast2GO [82].

Comparative genomics

Orthologous groups were constructed with OrthoMCL [83] using the proteome sequences of *B. tabaci* and 13 other insects, as well as two additional non-insect arthropod species (**Additional file 7**). Protein sequences of single-copy gene families were aligned with MUSCLE [84]. The resulting alignments were trimmed using trimAl [85] to remove positions with gaps in more than 20% of the sequences, and then used to reconstruct the phylogenetic tree using the Maximum-Likelihood method implemented in PhyML [86], the JTT model for amino acid substitutions and the aLRT method for branch support. Syntenic analysis between the five hemipteran genomes was performed using MCScanX [87].

A genome-wide screen for gene family expansions in the *B. tabaci* genome was performed based on InterPro domains. InterPro domains from the protein sequences of all the above 16 species were identified using InterProScan [81]. A domain was counted only once if it occurred multiple times in a protein sequence. Fisher's exact test was conducted for each domain, comparing the number of domains found in *B. tabaci* to the background, defined as the average of the counts in the other 15 species. The resulting raw p values were corrected for multiple testing using FDR [72]. An InterPro domain was considered to be significantly expanded in *B. tabaci* if the FDR was less than 0.05 and the count in *B. tabaci* was the largest among the 16 species in the comparison.

Symbiont genome assembly and annotation

Diagnostic PCR assays using the primers described in Pan et al. [43] indicated that the colony of MEAM1 used for genome sequencing bore the primary endosymbiont, Portiera, and two secondary endosymbionts, Hamiltonella and Rickettsia. Primers specific to Cardinium, Wolbachia, Fritschea, and Arsenophonus were also used in the whitefly enodymbiont screen but did not test positive. Primer sequences used were: Portiera F:5'-TGCAAGTCGAGCGGCATCAT-3', R:5'-AAAGTTCCCGCCTTATGCGT-3'; Rickettsia F:5'-GCTCAGAACGAACGCTATC-3', R:5'-GAAGGAAAGCATCTCTGC-3'; Hamiltonella F:5'-TGAGTAAAGTCTGGAATCTGG-3', R:5'-AGTTCAAGACCGCAACCTC-3'; Cardinium F:5'-GCGGTGTAAAATGAGCGTG-3', R:5'-ACCTMTTCTTAACTCAAGCCT-3'; Wolbachia F:5'-TGGTCCAATAAGTGATGAAGAAAC-3', R:5'-AAAAATTAAACGCTACTCCA-3'; Fritschea F:5'-GATGCCTTGGCATTGATAGGCGATGAAGGA-3', R:5'-TGGCTCATCATGCAAAAGGCA-3'; Arsenophonus F:5'-CGTTTGATGAATTCATAGTCAAA-3', R:5'-GGTCCTCCAGTTAGTGTTACCCAAC-3', which amplified approximately 1 kb, 0.9 kb, 0.7 kb, 0.4 kb, 0.6 kb, 0.6 kb, 0.6 kb of the respective gene [43]. PCR amplifications were performed in 20 ul reactions using GoTaq Green Master Mix (Promega, USA), 0.25 uM of each forward and reverse primer, and 150 ng DNA with initial denaturation at 95 C for 2 m, 30 cycles of denaturation at 95 ^{o}C for 30 s, annealing at 55 ^{o}C (Wolbachia), 57 C (Cardinium), 58 ^{o}C (Portiera, Hamiltonella, Arsenophonus), or 60 ^{o}C (Rickettsia, Fritschea) for 1 m, extension at 72 ^{o}C for 1 m, and a final extension at 72 ^{o}C for 5 m. PCR products were visualized on a 1% agarose gel.

The genomes of the three symbionts present in *B. tabaci*, i.e., *Portiera*, *Hamiltonella* and *Rickettsia*, were *de novo* assembled using the PacBio long reads. The error-corrected PacBio reads corresponding to the three symbiont genomes were first extracted by aligning the reads to the reference sequences of related species [88-90]. The extracted PacBio reads for each symbiont were *de novo* assembled using Sprai (http://zombie.cb.k.u-tokyo.ac.jp/sprai/). The final assembled contigs were corrected for base errors with iCORN2 [63] using the high-quality Illumina paired-end reads. Protein-coding genes from the three assembled genomes were predicted *ab initio* using GeneMark [91] and Glimmer [92]. The final consensus gene models were then derived using MAKER [75]. The predicted genes were functionally annotated by comparing their protein sequences against the UniProt database [93].

Identification of horizontal gene transfers

The *B. tabaci* genome sequences were first masked for repeat regions, and then translated in six frames. Potential polypeptides (PPPs) having lengths of at least 60 amino acids were kept. Furthermore, the high quality and cleaned RNA-Seq datasets were *de novo* assembled using Trinity [94]. The assembled contigs were aligned to the *B. tabaci*

genome and only those that could be aligned were used in the analysis. To identify HGTs of bacterial origin, the assembled transcript and genome-translated PPP sequences were compared against two protein databases derived from complete proteomes in UniProt [93], one consisting of eukaryotic proteins (excluding proteins from species in Arthropoda) and the other consisting of bacterial proteins. To identify HGTs of fungal origin, the assembled transcript and genome-translated PPP sequences were compared against the eukaryotic protein database (excluding proteins from species in Arthropoda and fungus) and the other consisting of fungus proteins. The index of horizontal gene transfer, h, was calculated by subtracting the bitscore of the best eukaryote match from that of the best bacteria/fungus match. We defined candidate HGTs as those with $h \geq 30$ and the bitscore of the best bacterial or fungus protein hit ≥ 100 as described in Crisp et al [46]. For each candidate HGT, we manually checked the alignments of DNA reads and RNA-Seq reads to genomic regions containing the HGT and the neighboring intrinsic insect genes, and provide the following evidence to support the HGT: 1) alignments of mate-pair DNA reads to support the assembly in regions containing the HGT and the neighboring insect genes; 2) coverage of paired-end DNA reads to support a HGT if the read depth of the HGT is similar to that of neighboring insect genes; 3) alignments of polyA enriched strand-specific RNA-Seq reads to support the structure and expression of the HGT. We then performed phylogenetic analysis to validate the bacterial or fungus origin of the HGTs. The protein sequence of each candidate HGT was compared against the protein databases of six taxa (archaea, bacteria, fungi, plants, metazoan, and other eukaryotes). The top five hits from each taxon were extracted, and aligned with the protein sequence of the candidate gene using ClustalW2 [95]. Each alignment was trimmed to exclude regions where gaps were more than 20% of sequences. Phylogenetic trees were constructed with PhyML [86] using a JTT model with 100 bootstraps. HGTs were considered validated if the genes were monophyletic with the bacterial or fungal taxa.

Acknowledgments

We thank Dr. Jim Giovannoni for critical reading of this manuscript, Drs. Ping Wang and Lei Gao for useful discussion, and Andrea Gilliard, April Bisner, Art Cortez and Laura Hladky for technical assistance; the Atkinson Center for a Sustainable Future (Cornell University) for financial support. This work was supported by grants from the USDA ARS Area-wide project to K.S.L. and W.M.W, the USDA-ARS Office of International Research Program from a grant provided by the USAID Feed-the-Future program (58-0210-3-012) to K.S.L., WMW and Z.F., NSF (IOS-1110080) to Z.F. and NSF (IOS-1109989 and IOS-1354309) to M.C., USDA NIFA (2016-67013-24756) to G.J. and A.E.D, Swedish Research Council U-forsk grant to M.C.S. and Israel Science Foundation (1127/13) to M.G.

Availability of data

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MAMS00000000. The version described in this paper is version MAMS01000000. Genome and transcriptome sequence reads have been deposited in the SRA as BioProjects PRJNA312470, PRJNA312467 and PRJNA347299, respectively. A whitefly genome database is available (http://www.whiteflygenomics.org).

Authors' contributions

Z.F., K.S.L., and W.M.W. designed the research and managed the project. D.K.H. and A.M.S. maintained the whitefly colony. D.K.H. isolated male individuals and prepared genomic DNA for sequencing. W.C. performed genome assembly, annotation, comparative genomic analysis, and RNA-Seq analysis. H.S. helped with genome assembly. D.K.H., N.K., A.Kliot, S.K. and G.L. performed the RNA-Seq experiments. Y.X. constructed RNA-Seq libraries. Y.Z., D.K.H. and N.K. contributed to RNA-Seq data analysis. A.E.D., W.C., J.L. and Z.F. contributed to the analysis of horizontally transferred genes. A.E.D., W.C., D.K.H., A.Kruse, M.C.S., D.R.N., G.J., J.K.B., J.L., M.C., M.G., N.K., P.V.P, T.W.F., W.B.H., X.Y., Y.L. contributed to annotation and analysis of specific gene families. W.L. and W.C. implemented the whitefly genome database. The authors declare no competing financial interests.

Competing interests

The authors declare that they have no competing interests.

Table 1 Summary of the *Bemisia tabaci* MEAM1 genome assembly

	Scaffold*	Contig*
Total number	19,762	52,037
Total sequences bp	615,077,135	
Maximum length	11,178,615	269,706
N50 length	3,232,964	29,920
L50 number	56	5,750
N90 length	381,346	6,117
L90 number	229	22,027
Gap length	14,380,491	0

*Only contigs and scaffolds >= 500 bp were included in the genome assembly

References

1. Martin J, Mound L: An annotated check list of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae). Lista de las moscas blancas del mundo (Insecta: Hemiptera: Aleyrodidae). *Zootaxa* 2007, 1(492):1-84.
2. Abd-Rabou S, Simmons AM: Survey of reproductive host plants of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Egypt, including new host records. *Entomol News* 2010, 121(5):456-465.
3. Navas-Castillo J, Fiallo-Olive E, Sanchez-Campos S: Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol* 2011, 49:219-248.
4. Scholthof K-b, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P: Top 10 plant viruses in molecular plant pathology. *Mol Plant Pathol* 2011, 12(9):938-954.
5. Lefeuvre P, Martin DP, Harkins G, Lemey P, Gray AJ, Meredith S, Lakay F, Monjane A, Lett J-M, Varsani A: The spread of *tomato yellow leaf curl virus* from the Middle East to the world. *PLoS Pathog* 2010, 6(10):e1001164.
6. Legg J, Jeremiah S, Obiero H, Maruthi M, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G: Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res* 2011, 159(2):161-170.
7. Legg JP, Shirima R, Tajebe LS, Guastella D, Boniface S, Jeremiah S, Nsami E, Chikoti P, Rapisarda C: Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. *Pest Manag Sci* 2014 70(10):1446-1453.
8. Patil BL, Legg JP, Kanju E, Fauquet CM: Cassava brown streak disease: a threat to food security in Africa. *J Gen Virol* 2015, 96(Pt 5):956-968.
9. Legg J, Somado EA, Barker I, Beach L, Ceballos H, Cuellar W, Elkhoury W, Gerling D, Helsen J, Hershey C: A global alliance declaring war on cassava viruses in Africa. *Food Security* 2014, 6(2):231-248.
10. Gennadius P: Disease of the tobacco plantations in the Trikonia. The aleurodid of tobacco. *Ellenike Georgia* 1889, 5:1-3.
11. Boykin LM: *Bemisia tabaci* nomenclature: lessons learned. *Pest Manag Sci* 2014, 70:1454-1459.
12. Brown JK: Bemisia: bionomics and management of a global pest: (eds Stansly PA & Naranjo S E) 31-67 (Springer, New York), 2010.
13. Alemandri V, Vaghi Medina CG, Dumon AD, Arguello Caro EB, Mattio MF, Garcia Medina S, Lopez Lambertini PM, Truol G: Three members of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) cryptic species complex occur sympatrically in argentine horticultural crops. *J Econ Entomol* 2015, 108(2):405-413.
14. Boykin LM, De Barro PJ: A practical guide to identifying members of the *Bemisia tabaci* species complex: and other morphologically identical species. *Front Ecol Evol* 2014,
15. Boykin LM, Armstrong KF, Kubatko L, De Barro PJ: Species delimitation and global biosecurity. *Evol Bioinformatics* 2012, 8:1–37.
16. De Barro PJ, Liu SS, Boykin LM, Dinsdale AB: *Bemisia tabaci*: A statement of species status. *Annu Rev Entomol* 2011, 56:1-19.

17. Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P: Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann Entomol Soc Am* 2010, 103:196–208.
18. De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB: *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol* 2011, 56:1-19.
19. Wang XW, Luan JB, Li JM, Su YL, Xia J, Liu SS: Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. *BMC Genomics* 2011, 12:458.
20. Wang XW, Zhao QY, Luan JB, Wang YJ, Yan GH, Liu SS: Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species. *BMC Genomics* 2012, 13:529.
21. Luan JB, Li JM, Varela N, Wang YL, Li FF, Bao YY, Zhang CX, Liu SS, Wang XW. Global analysis of the transcriptional response of whitefly to tomato yellow leaf curl China virus reveals the relationship of coevolved adaptations. *J Virol* 2011, 85:3330-3340.
22. Ilias A, Lagnel J, Kapantaidaki DE, Roditakis E, Tsigenopoulos CS, Vontas J, Tsagkarakou A: Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of *B. tabaci* reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. *BMC Genomics* 2015, 16:939.
23. Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS: De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* 2010, 11:400.
24. Luan JB, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling KS, Fei Z, Liu SS, Douglas AE: Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol* 2015, 7(9):2635-2647.
25. Ye XD, Su YL, Zhao QY, Xia WQ, Liu SS, Wang XW: Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. *BMC Genomics* 15:370.
26. Luan JB, Shan HW, Isermann P, Huang JH, Lammerding J, Liu SS, Douglas AE: Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. *Proc Biol Sci* 2016, 283(1833).
27. Shatters RG Jr, Powell CA, Boykin LM, Liansheng H, McKenzie CL: Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *J Econ Entomol* 2009, 102:750-758.
28. Chen W, Hasegawa DK, Arumuganathan K, Simmons AM, Wintermantel WM, Fei Z, Ling KS: Estimation of the whitefly *Bemisia tabaci* genome size based on k-mer and flow cytometric analyses. *Insects* 2015, 6(3):704-715.
29. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015, 31(19):3210-3212.
30. IAGC: Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol* 2010, 8(2):e1000313.

31. Consortium HG: Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 2012, 487(7405):94-98.
32. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG: Phylogenomics resolves the timing and pattern of insect evolution. *Science* 2014, 346(6210):763-767.
33. Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, Jiggins CD: Synteny and chromosome evolution in the Lepidoptera: evidence from mapping in Heliconius melpomene. *Genetics* 2007, 177(1):417-426.
34. Ghanim M: A review of the mechanisms and components that determine the transmission efficiency of *Tomato yellow leaf curl virus* (Geminiviridae; Begomovirus) by its whitefly vector. *Virus Res* 2014, 186:47-54.
35. Kubo Y, Hayashi H, Matsuyama T, Sato H, Yamamoto N: Retrovirus entry by endocytosis and cathepsin proteases. *Adv Virol* 2012, 2012:640894.
36. Sim S, Ramirez JL, Dimopoulos G: Dengue virus infection of the *Aedes aegypti* salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. *PLoS Pathog* 2012, 8(3):e1002631.
37. Horowitz AR, Ishaaya I: Dynamics of biotypes B and Q of the whitefly *Bemisia tabaci* and its impact on insecticide resistance. *Pest Manag Sci* 2014, 70(10):1568-1572.
38. Alon M, Elbaz M, Ben-Zvi MM, Feldmesser E, Vainstein A, Morin S: Insights into the transcriptomics of polyphagy: *Bemisia tabaci* adaptability to phenylpropanoids involves coordinated expression of defense and metabolic genes. *Insect Biochem Mol Biol* 2012, 42(4):251-263.
39. Hengst U, Albrecht H, Hess D, Monard D: The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors. *J Biol Chem* 2001, 276(1):535-540.
40. Vallee B, Coadou G, Labbe H, Sy D, Vovelle F, Schoentgen F: Peptides corresponding to the N-and C-terminal parts of PEBP are well-structured in solution: new insights into their possible interaction with partners in vivo. *J Pept Res* 2003, 61(2):47-57.
41. Corbit KC, Trakul N, Eves EM, Diaz B, Marshall M, Rosner MR: Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein. *J Biol Chem* 2003, 278(15):13061-13068.
42. Yeung KC, Rose DW, Dhillon AS, Yaros D, Gustafsson M, Chatterjee D, McFerran B, Wyche J, Kolch W, Sedivy JM: Raf kinase inhibitor protein interacts with NF-kappaB-inducing kinase and TAK1 and inhibits NF-kappaB activation. *Mol Cell Biol* 2001, 21(21):7207-7217.
43. Pan H, Li X, Ge D, Wang S, Wu Q, Xie W, Jiao X, Chu D, Liu B, Xu B *et al.*: Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. *PLoS ONE* 2012, 7: e30760.
44. Rao Q, Rollat-Farnier PA, Zhu DT, Santos-Garcia D, Silva FJ, Moya A, Latorre A, Klein CC, Vavre F, Sagot MF *et al*: Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. *BMC Genomics* 2015, 16(1):226.

45. Ranocha P, McNeil SD, Ziemak MJ, Li C, Tarczynski MC, Hanson AD: The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity. *Plant J* 2001, 25(5):575-584.
46. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G: Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol* 2015, 16(1):50.
47. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M: No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci USA* 2016, 113(18):5053-5058.
48. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC, Li Q, Jones CD, Yandell M *et al*.: Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci USA* 2015, 112(52):15976-15981.
49. Hotopp JCD, Clark ME, Oliveira DC, Foster JM, Fischer P, Muñoz Torres MC, Giebel JD, Kumar N, Ishmael N *et al.*: Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 2007, 317(5845):1753-1756.
50. Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in bdelloid rotifers. *Science* 2008, 320(5880):1210-1213.
51. Price DR, Wilson AC: A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol* 2014, 12:110.
52. Wollam J, Antebi A: Sterol regulation of metabolism, homeostasis, and development. *Annu Rev Biochem* 2011, 80:885-916.
53. Behmer ST, Nes WD: Insect sterol nutrition and physiology: a global overview. *Adv In Insect Phy* 2003, 31:1-72.
54. Behmer ST, Grebenok RJ, Douglas AE: Plant sterols and host plant suitability for a phloem-feeding insect. *Funct Ecol* 2011, 25(3):484-491.
55. Kluge M, Ullrich R, Dolge C, Scheibner K, Hofrichter M: Hydroxylation of naphthalene by aromatic peroxygenase from *Agrocybe aegerita* proceeds via oxygen transfer from H2O2 and intermediary epoxidation. *Appl Microbiol Biotechnol* 2009, 81(6):1071-1076.
56. Hammel KE, Cullen D: Role of fungal peroxidases in biological ligninolysis. *Curr Opin Plant Biol* 2008, 11(3):349-355.
57. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014, 30(15):2114-2120.
58. Kelley DR, Schatz MC, Salzberg SL: Quake: quality-aware detection and correction of sequencing errors. *Genome Biol* 2010, 11(11):R116.
59. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED: Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat Biotechnol* 2012, 30(7):693-700.
60. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H: Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* 2014, 24(8):1384-1395.

61. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al*: SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 2012, 1(1):18.
62. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC: Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 2012, 7(11):e47768.
63. Otto TD, Sanders M, Berriman M, Newbold C: Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* 2010, 26(14):1704-1707.
64. Kontsedalov S, Zchori-Fein E, Chiel E, Gottlieb Y, Inbar M, Ghanim M: The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. *Pest Manag Sci* 2008, 64(8):789-792.
65. Zhong S, Joung JG, Zheng Y, Chen YR, Liu B, Shao Y, Xiang JZ, Fei Z, Giovannoni JJ: High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* 2011, 2011(8):940-949.
66. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013, 41:D590-596.
67. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, 10(3):R25.
68. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, 25(9):1105-1111.
69. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010, 28(5):511-515.
70. Kim D, Langmead B, Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015, 12(4):357-360.
71. Robinson MD, McCarthy DJ, Smyth GK: EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26(1):139-140.
72. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995:289-300.
73. Bao Z, Eddy SR: Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res* 2002, 12(8):1269-1276.
74. Price AL, Jones NC, Pevzner PA: *De novo* identification of repeat families in large genomes. *Bioinformatics* 2005, 21 Suppl 1:i351-358.
75. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M: MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 2008, 18(1):188-196.
76. Han Y, Wessler SR: MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res* 2010:gkq862.
77. Stanke M, Waack S: Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 2003, 19 Suppl 2:ii215-225.

78. Korf I: Gene finding in novel genomes. *BMC bioinformatics* 2004, 5:59.
79. Gotoh O: Direct mapping and alignment of protein sequences onto genomic sequence. *Bioinformatics* 2008, 24(21):2438-2444.
80. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S *et al*: The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 2015, 43:D213-221.
81. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al*: InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014, 30(9):1236-1240.
82. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, 21(18):3674-3676.
83. Li L, Stoeckert CJ, Jr., Roos DS: OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003, 13(9):2178-2189.
84. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, 32(5):1792-1797.
85. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009, 25(15):1972-1973.
86. Guindon S, Delsuc F, Dufayard JF, Gascuel O: Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol* 2009, 537:113-137.
87. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H: MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012, 40(7):e49-e49.
88. Rao Q, Wang S, Su YL, Bing XL, Liu SS, Wang XW: Draft genome sequence of "*Candidatus Hamiltonella defensa*," an endosymbiont of the whitefly *Bemisia tabaci*. *J Bacteriol* 2012, 194(13):3558.
89. Rao Q, Wang S, Zhu DT, Wang XW, Liu SS: Draft genome sequence of *Rickettsia* sp. strain MEAM1, isolated from the whitefly *Bemisia tabaci*. *J Bacteriol* 2012, 194(17):4741-4742.
90. Jiang ZF, Xia F, Johnson KW, Brown CD, Bartom E, Tuteja JH, Stevens R, Grossman RL, Brumin M, White KP *et al*: Comparison of the genome sequences of "*Candidatus Portiera aleyrodidarum*" primary endosymbionts of the whitefly *Bemisia tabaci* B and Q biotypes. *Appl Environ Microbiol* 2013, 79(5):1757-1759.
91. Besemer J, Borodovsky M: GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 2005, 33(Web Server issue):W451-454.
92. Salzberg SL, Delcher AL, Kasif S, White O: Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 1998, 26(2):544-548.
93. UniProtConsortium: UniProt: a hub for protein information. *Nucleic Acids Res* 2015, 43:D204-212.
94. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, 29(7):644-652.

95. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, 23(21):2947-2948.

Figure legends

Figure 1. **Whitefly phylogenomics and gene family expansions. (A)** Adult whiteflies *Bemisia tabaci*. ♀, female; ♂, male. (**B)** Phylogenetic relationship and gene orthology of *B. tabaci* and other arthropods. 1:1:1 indicates single-copy genes in all species; N:N:N indicates multi-copy genes in all species; Insect specific refers to genes present only in the 14 insect species; Endopterygota refers to genes present only in at least two endopterygotan insects; Exopterygota refers to genes present only in at least two exopterygotan insects. (**C)** Significantly expanded domains in *B. tabaci*. Bta, *B. tabaci*; Api, *Acyrthosiphon pisum*; Dci, *Diaphorina citri*; Rpr, *Rhodnius prolixus*; Nlu, *Nilaparvata lugens*; Phu, *Pediculus humanus*; Nvi, *Nasonia vitripennis*; Ame, *Apis mellifera*; Cfl, *Camponotus floridanus*; Tca, *Tribolium castaneum*; Bmo, *Bombyx mori*; Dpl, *Danaus plexippus*; Aga, *Anopheles gambiae*; Dme, *Drosophila melanogaster*; Dpu, *Daphnia pulex*; Tur, *Tetranychus urticae*.

Figure 2. **Whitefly genes associated with virus acquisition and transmission.** (**A)** Heatmap of differentially expressed cathepsin genes (cts) in whiteflies upon acquisition of TYLCV or ToCV after 24, 48 and 72 hours, respectively. Three biological replicates were performed for each sample. Color indicates fold change of gene expression (viruliferous/non-viruliferous whiteflies). (**B)** Phylogenetic tree of cysteine proteinase-type cathepsins in *B. tabaci* and other species. Maximum likelihood tree was constructed using amino acid sequences of the peptidase C1A domain. _HUMAN, cathepsins from human; _MOUSE, cathepsins from mouse; _RAT, cathepsins from rat; _DROME, cathepsins from fruit fly. (**C)** Gene clusters containing whitefly-specific genes that were differentially expressed upon ToCV acquisition. Genes marked by asterisk are differentially expressed. Genes in same colors in each cluster are duplicated genes while genes in white are non-duplicated

Figure 3. **Whitefly genes responsive to insecticide Mospilan.** (**A)** Number of Mospilan-responsive genes encoding phosphatidylethanolamine-binding protein (PEBP), cathepsin (CTS), cytochrome P450 (CYP), UDP-glucuronosyltransferase (UGT), carboxylesterase (CCE), ABC transporter (ABC), glutathione S-transferase (GST) and aromatic peroxygenase (APO) in susceptible and resistant MED populations. Three or four biological replicates were performed for each control or Mospilan-treated sample. Detailed expression information of these genes is provided in Additional file 14. (**B)** Phylogenetic tree of PEBPs in *B. tabaci* and other arthropod species. (**C)** Two largest clusters of PEBPs on Scaffold1195, with 34 and 38 copies, respectively. Red, PEBP genes located in the positive strand of the scaffold; Blue, PEBP genes located in the negative strand of the scaffold; white, non-PEBP genes; Genes marked with asterisk are Mospilan-responsive.

Figure 4. **Horizontal gene transfers in whitefly.** (**A)** Genome synteny of *panB-panC* between *B. tabaci* and *Pseudomonas*. Red, positive strand; blue, negative strand; pink, untranslated regions. **(B)** Coordinated pathway of the pantothenate synthesis between *B. tabaci* and its symbiont *Portiera*. Gene in blue is *B. tabaci* intrinsic, gene in orange is from *Portiera*, and genes in red are horizontally transferred. (**C**) Phylogenetic tree of *B. tabaci* squalene-hopene cyclases of bacterial origin and those from other kingdoms. Numbers on branches represent bootstrap values, only those >90 are shown.

Additional files

Additional file 1: Supplementary text. **Figure S1**. Whitefly (*Bemisia tabaci* MEAM1 or B biotype) life cycle. **Figure S2**. Genome clusters containing whitefly-specific unknown genes that are differentially expressed upon ToCV acquisition. **Figure S3**. Phylogenetic tree of cytochrome P450s from *Bemisia tabaci* and other species. **Figure S4**. Phylogenetic tree of GST family genes from *Bemisia tabaci* and other species. **Figure S5**. Phylogenetic tree of ABC transporters. **Figure S6**. Large tandem clusters of PEBP (phosphatidylethanolamine-binding protein) genes in the *Bemisia tabaci* genome. **Figure S7**. Circular view of the genomes of *Bemisia tabaci* endosymbionts. **Figure S8**. Amino acid biosynthesis pathways in *Bemisia tabaci* and its endosymbiont bacteria. **Figure S9**. Validation of HGTs using mate-pair and paired-end genome reads, and RNA-Seq reads. **Figure S10**. Genome synteny of bioA-bioD between *Bemisia tabaci* and *Cardinium*. **Figure S11**. Number of immunity-related genes across various insect species. **Figure S12**. Phylogenetic analysis of *Bemisia tabaci* RNA-dependent RNA polymerases.

Additional file 2: Summary of *Bemisia tabaci* genome sequencing data.

Additional file 3. Summary of RNA-Seq dataset.

Additional file 4. Mapping statistics of *Bemisia tabaci* mRNA sequences to the *B. tabaci* genome.

Additional file 5. Repeat sequences in the *Bemisia tabaci* genome assembly.

Additional file 6. Statistics of functional annotation of *Bemisia tabaci* predicted genes.

Additional file 7. *Bemisia tabaci* genome annotation and comparison with fruit fly.

Additional file 8. Source of Proteomes used for comparative genomics analysis.

Additional file 9. Domain count of expanded families in *Bemisia tabaci*.

Additional file 10. Differentially expressed genes in *Bemisia tabaci* after acquisition of TYLCV or ToCV for 24, 48, and 72 hours, respectively.

Additional file 11. Number of cathepsin genes in *Bemisia tabaci* as compared to other arthropods.

Additional file 12. Number of genes potentially involved in detoxification and insecticide resistance.

Additional file 13. Cytochrome P450 genes in *Bemisia tabaci*.

Additional file 14. Genes from detoxification and other interesting families that are differentially expressed upon insecticide treatment.

Additional file 15. Endosymbiont genome assembly and annotation.

Additional file 16. Amino acid biosynthesis pathway in *B. tabaci* and its endosymbionts.

Additional file 17. Horizontally transferred genes of bacterial origin in *Bemisia tabaci*.

Additional file 18. Horizontally transferred genes of fungal origin in *Bemisia tabaci*.

Additional file 19. Immunity-related genes in *Bemisia tabaci*.

Additional file 20. List of genes in the miRNA and siRNA pathway in the *Bemisia tabaci* genome.